Doctoral theses at NTNU, 2018:132

Zhe Li

Deep Learning Driven Approaches for Predictive Maintenance

A Framework of Intelligent Fault Diagnosis and Prognosis in the Industry 4.0 Era

Doctoral thesis

Zhe Li

ISBN 978-82-326-3054-7 (printed ver.)
ISBN 978-82-326-3055-4 (electronic ver.)
ISSN 1503-8181

Doctoral theses at NTNU, 2018:132

NTNU
Norwegian University of Science and Technology
Thesis for the Degree of Philosophiae Doctor
Faculty of Engineering
Department of Mechanical and Industrial Engineering

NTNU
Norwegian University of Science and Technology

NTNU
Norwegian University of Science and Technology

NTNU

Zhe Li

Deep Learning Driven Approaches for Predictive Maintenance

A Framework of Intelligent Fault Diagnosis and Prognosis in the Industry 4.0 Era

Thesis for the Degree of Philosophiae Doctor

Trondheim, April 2018

Norwegian University of Science and Technology
Faculty of Engineering
Department of Mechanical and Industrial Engineering

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the Degree of Philosophiae Doctor

Faculty of Engineering
Department of Mechanical and Industrial Engineering

ISBN 978-82-326-3054-7 (printed ver.)
ISBN 978-82-326-3055-4 (electronic ver.)
ISSN 1503-8181

Doctoral theses at NTNU, 2018:132

Printed by NTNU Grafisk senter

Acknowledgements

It gives me an immense pleasure to present the thesis in its completed form. There have been a number of persons that helped in this work either on the professional, personal, or both. First and foremost, I would like to express my sincere gratitude to my main supervisor, Professor Ph.D. Kesheng Wang for his guidance and constant support throughout the Ph.D. study. Without his timely advice and knowledge in data mining, I would never have accomplished the research work.

I would like to thank my co-supervisor Professor Jan Ola Strandhagen for his support and help during these years.

I thank my co-supervisor Professor Yafei He from Shanghai Polytechnic University. His support pushed me forward during the study.

I wish to thank the administrative staff of the Department of Mechanical and Industrial Engineering for providing excellent working facilities and environment.

I also would like to thank my colleagues Postdoc Quan Yu, Ph.D. Haishu Ma, Lecturer Zhiyong Cai, Ph.D. Candidate Lelija Stupar and Postdoc Jin Yuan, for their supports and helps during the Ph.D. study.

I also wish to thank Associate Professor Zongzheng Ma, Associate Professor Shewei Wang, and Ms. Xiaodan Luo, who came to NTNU as visiting scholars. It is my pleasure to work with them during their visiting.

Special thanks go to Lecturer, PhD Yi Wang from Plymouth University for his guidance and assistance during the study.

Last but not least, I am extremely grateful to my parents for being there all these years, for their patience, encouragement and support.

Trondheim, March 2018

Zhe Li

Abstract

Nowadays, with the trend of smart manufacturing and development of information and communication technologies, companies are increasingly applying types of sensors and information technologies to capture data at all stages of production. Simultaneously, technologies such as Internet of Things (IoT), Internet of Services (IoS), Artificial intelligence (AI), and data mining (DM), which are all inherent in Industry 4.0, are being leveraged with "Big Data" to facilitate a more adaptable and smart maintenance policy.

Predictive maintenance is a type of maintenance policy raised under this background. The goal of predictive maintenance is to reduce the downtime and the cost of maintenance under the premise of zero failure manufacturing through utilizing real-time data to forecast potential faults. The philosophy of predictive maintenance is to perform maintenance only when necessary, which means maintenance shall only take place after analytical models forecast certain failures or degradations. Ideally, maintenance schedule can be optimized to minimize the cost of maintenance and achieve zero failure manufacturing through this policy. However, it is difficult to realize all the advantages of predictive maintenance without the foundations of correlation techniques.

This thesis presents a framework driven by deep learning approaches for predictive maintenance concerning the Industry 4.0 concept. This framework aims to provide an overall understanding and helpful guidance for researchers and practitioners to implement predictive maintenance in the Industry 4.0 era. The target of the framework is to minimize the number of unnecessary maintenance performance, leverage the remaining useful life of equipment, and reach zero failure manufacturing through intelligent diagnosis and prognosis. The framework consists of three tiers. The first tier is data acquisition from multiple sources, which allows all relevant devices and data sources interconnect with each other. Diagnosis and prognosis is the second tier, which is responsible to deal with all DM and analysis issues. The third tier is decision support and maintenance implementation. It is in charge of presenting all the information or knowledge obtained from DM, providing optimized maintenance scheduling and finally interacting with the physical world according to the computation results in cyber world.

The outcomes of the thesis can be applied in mechanical and electrical system in industries of manufacturing.

Table of Contents

List of Figures

List of Tables

Nomenclature

AI	Artificial Intelligence
AEA	Acoustic Emission Analysis
ANNs	Artificial Neural Networks
BP	Back Propagation
BPNN	Back Propagation Neural Network
BPSO	Binary Particle Swarm Optimization
CCNN	Cascade Correlation Neural Network
CI	Computational Intelligence
CNC	Computer Numerical Control
CNN	Convolutional Neural Networks
CPNN	Counter Propagation Neural Networks
CPS	Cyber-Physical Systems
DBN	Deep Belief Network
DM	Data Mining
DNN	Deep Neural Network
DT	Decision Tree
DWNN	Dynamic Wavelet Neural Network
DWT	Discrete Wavelet Transform
EMD	Empirical Mode Decomposition
ES	Expert System
FFT	Fast Fourier Transform
FLS	Fuzzy Logic System
GPU	Graphics Processing Unit
HDPS	Hierarchical Diagnosis and Prognosis System
HHT	Hilbert-Huang Transform
ICT	Information and Communication Technologies

IoS	Internet of Services
IoT	Internet of Things
KDD	Knowledge Discovery from Data
KNNC	K-Nearest Neighbours Classification
KPI	Key Performance Indicators
LSTM	Long Short Term Memory
MLP	Multi-layer Perception
MSE	Mean Squared Error
MSO	Maintenance Scheduling Optimization
NFN	Neuro-Fuzzy Network
OA	Oil Analysis
PM	Predictive Maintenance
PNN	Polynomial Neural Network
PSO	Particle Swarm Optimization
RBFNN	Radial Basis Function Neural Networks
RBM	Restricted Boltzmann Machines
RFID	Radio Frequency Identification
RNN	Recurrent Neural Networks
SAE	Stacked Autoencoders
SCADA	Supervisory Control and Data Acquisition
SOM	Self-organising Map
STFT	Short Time Fourier Transform
SVM	Support Vector Machines
SVMR	Support Vector Machine Regression
TA	Temperature Analysis
VA	Vibration Analysis
WD	Wavelet Decomposition
WPD	Wavelet Packet Decomposition

Chapter 1 Introduction

1.1 Motivation

Nowadays, due to the rapid development of modern manufacturing industry, and information and communication technologies (ICT), engineering systems are increasingly becoming complex and integrated, which means unanticipated faults could result in consequences that range from the simple replacement of a cheap bearing to an accident that may cost millions in lost production, human sources, or pollution. Therefore, conventional maintenance strategies, which do not have the ability to completely eradicate faults, may no longer fulfil the requirement of modern industry. Actually, how to reduce downtime and cost of maintenance under the premise of zero failure manufacturing is always a critical issue for a company to be competitive and sustainable.

As an ideal maintenance policy, predictive maintenance collects types of parameters from equipment, detects changes in the physical condition of equipment, and discovers fault information, including when, where, and which type of fault that may occur. Through the fault information, predictive maintenance could arrange appropriate maintenance performance for maximizing the service life of equipment without increasing the risk of failure. Prediction for future potential fault allows enough time for maintenance planning before the fault happens. Ideally, maintenance schedule can be optimized to minimize the cost of maintenance and achieve zero failure manufacturing.

The key to implement predictive maintenance is ability to assess equipment health and discover detail information about current or future faults through collected data. Normally, it can be divided into two stages. One is so-called fault diagnosis, which means detecting, isolating, and identifying incipient failures or current degradations of certain components and performance. The other is fault prognosis, whose task is to identify and predict the impending or potential failures or degradations for the equipment, and subsequent monitoring and tracking for the growth of these failures. For decades, to construct a machine learning system for fault diagnosis and prognosis requires elaborate engineering and considerable domain expertise. Actually, very few research involves accurate prediction of certain potential failures for equipment, because of the challenge to precisely forecast the temporal progression of potential or impending faults. Actually, foretelling the future in a wide field of disciplines from engineering science, biology and economics to geography or sociology has attracted the interest of both researchers and practitioners over the past few decades, with results that vary from disappointing to promising.

From the author's perspective, the main obstacles to implement predictive maintenance could be concluded into three points: (1) Access to obtain necessary and massive industrial data, which can represent working condition for equipment. (2) Capability to integrate and leverage industrial big data for fault diagnosis and prognosis (3) Ideal data-driven models, which can accurately predict the potential or impending faults. Fortunately, with the rapid development of ICT and artificial intelligence (AI) techniques, along with the trend of Industry 4.0 and deep learning approaches, predictive maintenance is also on the threshold of a new era. Therefore, during the three years of PhD work, a framework for predictive maintenance concerning the Industry 4.0 concept and deep learning-driven approaches is established. The framework could offer a complete understanding and effective guidance for researchers and practitioners to implement predictive maintenance in the new situation. In addition, an overall research about the application of deep learning approaches, which could ideally get rid of some 'bottlenecks' that conventional data-driven methods faced in predictive maintenance along with detailed case studies are also presented in this thesis.

1.2 Contributions

This thesis was conceived with the objective of pushing forward research on predictive maintenance in Industry 4.0 concept. The main original contributions of this monograph are the following:

- A detailed revision of the state-of-the-art, the most relevant developments made in the field of predictive maintenance. Effort was put into producing a comprehensive survey and review as a starting point for new researchers into the field, and providing current researchers or practitioner a broader overview of all the approaches applied to achieve predictive maintenance with Industry 4.0 concept.
- A systematic research of deep learning based fault identification and prediction. Several deep learning architectures have been investigated from practical applications to interpret their superiorities in fault identification and prediction in

certain domains or with prerequisites, which could provide effective guidance to select suitable deep learning methods to implement predictive maintenance. Some novel applications of deep learning-driven fault diagnosis and prognosis approaches such as DNN-based degradation assessment, DBN-based error prediction, SAE-based feature reconstruction, and LSTM-based anomaly detection are also presented along with detailed case studies in the thesis, which demonstrate the superiorities of deep learning in self-learning, big data analysis, fault identification, and degradation assessment.

- The framework for predictive maintenance concerning the Industry 4.0 is proposed in the thesis to achieve accurate failure prediction, efficient maintenance scheduling, extension of data sources and tracking degradation of equipment. As a result, the efficiency of maintenance implementation could be enhanced. The framework also provides an overall understanding and helpful guidance for researchers and practitioners to implement predictive maintenance in the Industry 4.0 era.
- An Industrial 4.0 scenario about the implementation of predictive maintenance for machining centers together with a case study of DBN-based backlash error prediction are demonstrated in the thesis. In the case study, a novel HDPS-BPSO maintenance implementation strategy is proposed to achieve predictive maintenance in practical applications. The numerical result in that case not only proves the superiority of deep learning methods in knowledge discovery with strong self-learning ability but also demonstrates the benefit of implementing predictive maintenance compared with preventive maintenance.
- An experiment of fault classification and degradation assessment for rotary machinery is also presented in the thesis to provide a comprehensive comparison of different types of data driven models, in which DNN-based degradation assessment outperforms the other models and proves the advantages of deep learning. In this case, a novel SAE-LSTM approach is also proposed for anomaly detection to train data-driven models in an unsupervised learning environment when the empirical knowledge is missing and all data is collected without labels.

1.3 List of scientific articles

- Zhe Li, Yi Wang, and Kesheng Wang. (2017) Applying Data-driven Method Based on Deep Belief Networks to Predict Backlash Error in a Machining center. *Journal of Intelligent Manufacturing*. DOI: 10.1007/s10845-017-1380-9.
- Zhe Li, Yi Wang, and Kesheng Wang. (2017) Intelligent Predictive Maintenance for Fault Diagnosis and Prognosis in Machine Centers - Industry 4.0 scenario. *Advances in Manufacturing*. https://doi.org/10.1007/s40436-017-0203-8.
- Zhe Li, Yi Wang, and Kesheng Wang. (2017) A Deep Learning Driven Method for Fault Classification and Degradation Assessment in Mechanical Equipment. (*Submitted to Computers in Industry*)
- Zhe Li, Yi Wang, Kesheng Wang, and Jin Yuan. (2018) A Deep Learning Approach for Anomaly Detection based on SAE and LSTM in Mechanical Equipment. (*Submitted to Expert System with Application*)
- Kesheng Wang, Zhe Li, Jørgen Braaten, and Quan Yu. (2015) Interpretation and Compensation of Backlash Error Data in Machine Centers for Intelligent Predictive Maintenance Using ANNs. *Advances in Manufacturing*, 3(2), 97-104.
- Zhe Li and Kesheng Wang. (2015) Comparing Different Methods of Fault Classification in Centrifugal Pumps. *WHITE Transactions on Engineering Sciences* 113 (2016): 217-224.
- Zhe Li and Kesheng Wang. (2015) Research on Fault Diagnosis and Prognosis for Machine Centers. *Advanced Manufacturing and Automation V*.
- Zhe Li, Kesheng Wang, and Yafei He. (2016) Industry 4.0 - Potentials for Predictive Maintenance. *International Workshop of Advanced Manufacturing and Automation* 2016.
- Zhe Li, Kesheng Wang, and Yavor Stefanov. (2016) Applying Radial Basis Function Networks to Fault Diagnosis of Motorized Spindle. *International Workshop of Advanced Manufacturing and Automation* 2016.

- Jinghui Yang, Yavor Stefanov, and Zhe Li. (2016) Applying Built-in Virtual Personal Assistant for Educational Equipment. *Atlantis Press. 2016. ISBN 978-94-6252-243-5. Advances in Economics, Business and Management Research (IWAMA-16)*.
- Shewei Wang, Kesheng Wang, and Zhe Li. (2016) A Review on Data-driven Predictive Maintenance Approach for Hydro Turbines/Generators. *Atlantis Press. 2016. ISBN 978-94-6252-243-5. Advances in Economics, Business and Management Research (IWAMA-16)*.
- Zhe Li, Viggo Gabriel Borg Pedersen, Kesheng Wang, and Yafei He. (2017) Fault Classification and Degradation Assessment Based on Wavelet Packet Decomposition for Rotary Machinery. *International Workshop of Advanced Manufacturing and Automation* 2017.

1.4 Outline of thesis

The thesis is structured in seven chapters. Chapter 1 introduces the motivation and contribution of this thesis. Chapter 2 presents a review of literature relevant to development of maintenance policy, predictive maintenance, impact of Industry 4.0, and application of deep learning. Chapter 3 describes the general structure of proposed framework to implement predictive maintenance concerning the Industry 4.0 concept and deep learning approaches. Chapter 4 provides helpful guidelines to select suitable fault analysis techniques and implement predictive maintenance in machining centers. Chapter 5 introduced a case study of implementing predictive maintenance for backlash error compensation. Chapter 6 demonstrates an experiment of fault classification, degradation assessment, and anomaly detection for rotary machinery through deep learning approaches. Chapter 7 concludes the thesis and proposes the further research.

Chapter 2
Literature review

2.1 Introduction

The requirements of repair for equipment, even the most rudimentary tools, exist from the very beginning of human civilization. In that days, human would do maintenance only when the equipment is "breakdown", which means it is no longer possible to work. This kind of maintenance is called as corrective or breakdown maintenance today. It was the only maintenance strategy until around 1950, some Japanese engineers proposed a new concept in maintenance, to lubricate and replace certain parts of equipment before it broke down. The new strategy was called as "Preventive Maintenance". From then on, plant managers were encouraged to develop programs for lubricating or replacing parts to protect the equipment from breakdown. Although it is useful and helps to reduce downtime, it is an expensive alternative, since many parts were replaced on a time-basis, while they could have lasted longer.

Simultaneously, with the development of modern manufacturing technology, new engineering systems are becoming more and more complex and integrated. Various components and subsystems may work together to accomplish certain missions. When a fault occurs, it is critical to identify the consequences and causes as rapidly as possible, and take appropriate maintenance action. Typically, when a system goes down, only a small fraction of the downtime can be spent to detect the root cause that leads to the fault. An unexpected failure may result in a devastating accident and financial losses for the company. Consequently, the ability for early prediction, which can prevent failures from growing and eventually turning into serious problems, is meaningful and imperative for industrial scenarios [Henriquez et al., 2014]. Therefore, in recent years, some researchers proposed a new type of maintenance, predictive maintenance. The idea is to schedule maintenance according to the available information, which can indicate the current condition of the equipment, or predict certain degradations. However, it was hard to achieve this vision because of technological capability and difficulties to acquire and integrate all essential information in the last few decades. Nowadays, with the trend of the fourth industrial revolution, we also see the potentials and challenges of predictive maintenance in this new era.

This chapter reviews the development of predictive maintenance, impact of Industry 4.0, and trend of deep learning approaches related to maintenance strategy. Various research, applications, and methods for faults diagnosis and prognosis are also summarized.

2.2 Impact of Industry 4.0

Industrial production keeps stepping forward since the very beginning. Sometimes, the changes were so potent and significant that we have to describe them with the term "Industrial Revolutions". The term, Industry 4.0, is used to recognize the other three previous industrial revolutions.

The first industrial revolution is the term used to describe the change from purely manual work to machine production, which initially affected the cotton-spinning and weaving mills in England from 1770. The great breakthrough came in 1782 with the steam engine invented by James Watt. Since then it was feasible to acquire energy supply at any location and any time, and the manual work was no longer focused as before.

The second industrial revolution was characterized by the principles of rationalization by Taylor. It is mainly based on the division of labour, precision manufacturing, standardization, and assembly line work. Henry Ford applied the first conveyor belt in the production of the T-model and achieved pioneering and great success with it in the automobile manufacturing at the beginning of the 20^{th} century.

The third industrial revolution focused on the developments of the computer and IT technology. This revolution led to numerically controlled machines, such as numerical control machines and industrial robots, which could be modified much faster and more efficient than conventional mechanical automated machines. Thus, the term of flexible automation was born and systems could be characterized by high productivity and flexibility.

Today, we are witnessing the fourth industrial revolution, which also known as Industry 4.0. It combines strengths of optimized industrial manufacturing with internet technologies, cyber-physical systems (CPS), internet of things (IoT) and internet of services (IoS) [Lasi et al., 2014], which changes manufacturing process, maintenance management and maintenance strategies significantly. Information and communication technologies are growing together and affecting all areas of life. Devices and systems in our real environment that are controlled by embedded software are now integrated into the global communication network, where ‘internet’ is the key term. The real world and the virtual world are clearly growing together. As a kind of buzzwords today, Industry 4.0 is widely discussed among practitioners as well as theorists, and facilitates the version of smart factory [Zamfirescu et al., 2013]. It was introduced at Hanover fair in 2011 in Germany to present a new trend towards the networking of traditionally industries [K. Wang, 2016]. From then on, there are many similar projects and programs demonstrated to reflect the concept of Industry 4.0 such as “Intelligent Manufacturing system”, and “Smart Manufacturing”.

Industry 4.0 is closely related to other technological concepts, such as Machine-to-Machine (M2M) communication [Gorecky et al., 2014], radio frequency identification (RFID) technology [K. Wang, 2014], CPS [K. Wang and Wang, 2012], the Internet of Things , the Internet of Services , cloud computing [Drath and Horch, 2014], computational intelligence (CI), data mining and decision-making/supporting system. In an Industry 4.0 factory, machines are connected as a collaborative community to collect, exchange and analyse data systematically. It combines strengths of optimized industrial manufacturing with internet technologies and changes manufacturing process, maintenance strategies and maintenance management significantly. Therefore, many companies face the challenge to assess the diversity of developments and concepts summarized the term Industry 4.0 and to develop their own strategies [Hochschild, 2015]. However, since lack of research of the potential use of Industry 4.0 and prospect of predictive maintenance, many companies and organizations are exposed to a dilemma, neither to wait too long with their Industry 4.0 implementation nor to start too early and commit fatal errors [Schmidt et al., 2015]. Therefore, this chapter aims to provide empirical knowledge on the potentials of Industry 4.0 from the perspective of maintenance. It may help academics and practitioners to identify and prioritize their steps towards predictive maintenance and condition-based maintenance management under the environment of Industry 4.0.

Based on the literature review, the general definition of Industry 4.0 can be summarized as that Industry 4.0 is a collective term for technologies and concepts of value chain organization [Thuemmler and Bai, 2017]. The core principle of Industry 4.0 is the application of IoT and smart manufacturing, which makes components and production machines collect and share data in real time, and take the best advantage of these expensive resources [Shrouf et al., 2014]. Within the smart factories of Industry 4.0, CPS monitor physical processes, create a virtual copy of the physical world and hereby make decentralized decisions. Over the IoT, CPS communicate and cooperate with each other and humans in real time. Data mining (DM) discover knowledge to support decision-making process. Through IoS, both internal and cross-organizational services are offered and utilized by participants of the value chain [K. Wang, 2016].

Industry 4.0 is the superposition of several technological developments related to CPS, IoT, IoS and DM. CPS refers to a new generation of systems with integrated computational and physical capabilities that can interact with humans through many new modalities. The key is the ability to interact with, and expand the capabilities of the physical world through computation, communication, and control [Baheti and Gill, 2011]. In addition, Industry 4.0 facilitates the development of intelligent and flexible production control systems, which apply information and communication technologies to make machines have the ability to intercommunicate and interact. Typically, Industry 4.0 mainly consists of following key components:

- Cyber-Physical Systems;
- Internet of Things;
- Big Data & Data Mining;
- Internet of Service.

Figure 2.1 shows the roles of these key components. As an important component of Industry 4.0, CPS refers to a new generation of systems with integrated computational and physical capabilities that can interact with humans through many new modalities. The target is to bring the virtual and physical worlds together to create a truly networked world in which intelligent objects communicate and interact with each other [MacDougall, 2014]. The key is the ability to interact with, and expand the capabilities of, the physical world through computation, communication, and control [Baheti and Gill, 2011]. CPS has the ability to transfer the physical world into the virtual one and can be understood as a basic unit. The development and application of identification approaches such as radio-frequency identification has become the foundation to achieve unique identification of objects. CPS applies multiple sensors with information and communication technologies to collect, store, and parse data. CPS deeply embeds cyber capabilities in the physical world, either on humans, infrastructure or platforms, to transform interactions with the physical world [Poovendran, 2010]. Therefore, CPS can be considered as the architecture in Industry 4.0 concept.

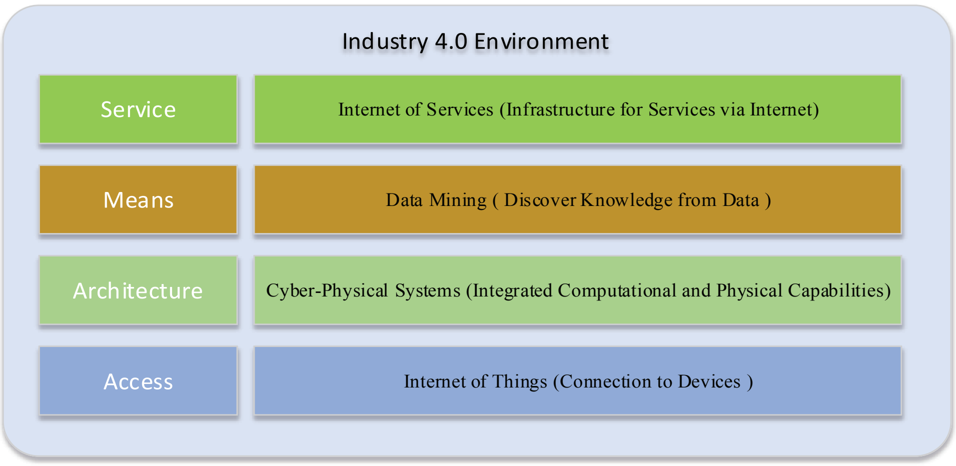

Figure 2.1 Key components in Industry 4.0

IoT is defined as the ubiquitous access to entities on the internet for the extension of the physical world through a variety of sensing, detection, identification, location tracking and monitoring equipment [Chaves and Nochta, 2011]. It allows "Things or Objects" interact with each other and cooperate with their 'smart' components to reach common aims. IoT can be thought as a network where CPS cooperates with each other through unique addressing schemas. It is the infrastructure that enables the internal connection of all types of devices through the internet for data sharing and information publication. Therefore, we can take the best advantage of collected data or information.

DM can be defined as the process of discovering interesting (non-trivial, implicit, previously unknown and potentially useful) patterns and knowledge from large amounts of data [Arslan et al., 2016]. It can also be considered as a result of the natural evolution in information technology. Actually, the evolution is an essential process, where intelligent methods are applied to extract data patterns and discover knowledge from data [Sohrabi and Akbari, 2016]. The data sources may include databases, data warehouses, the Web, other information repositories, or data that are streamed into system dynamically [Han et al., 2011]. In one word, data mining is the process of answering questions by searching a database for rules, relationships and patterns not found by conventional query tools.

IoS pursuits a similar approach with services instead of physical entities. The integration of these developments promotes the cooperation between the partners alone the entire system. It enables service vendors to offer their services via the internet. The IoS consists of business models, an infrastructure for services, the services themselves and participants. Services are offered and combined into value-added services by various suppliers. They are communicated to users as well as consumers and accessed by them via various channels.

Industry 4.0 also facilitate the vision and execution of the idea "Smart Factory", in which CPS monitor physical processes, create the virtual copy to represent the physical world and make decentralized decisions. Products may not only provide their identity but also record their properties, history and status via ICT technology. Over cloud computing and distributed control, CPS communicate and cooperate with each and via the Internet of Services, both internal and cross-organizational services are offered and utilized to make machines self-aware and actively prevents potential performance issues [Hermann et al., 2015].

From the perspective of maintenance, a self-aware and self-maintained machine system can be considered as a system which can self-assess its own health and degradation, and further use similar information from other peers for smart maintenance decisions to avoid potential faults [Lee et al., 2014]. To achieve such intelligence, smart analytics may be used at the individual machine or fleet levels. For a mechanical system, self-aware means the capability to assess the current, past or future working condition of a machine, and output the evaluation result. Such health assessment can be performed through data mining technologies to analyse the information collected from the given machine and its ambient environment. In this situation, real-time big data is no longer just a process for storing a huge amount of data in a data base or warehouse. DM enables us to analyse and discover patterns, rules and knowledge from big data collected from multiple sources. Therefore, we can make the most appropriate decision at the right time and right place according to the analysis result from real-time data. However, there still remains a gap between achieving all these visions and systematic research to guide the implementation of these intelligent maintenance in practical applications. For this reason, a framework to implement predictive maintenance in the Industry 4.0 era will be proposed in Chapter 3 to provide an overall understanding and helpful guidance for researchers and practitioners.

2.3 Maintenance strategy classification

According to European standard EN 13306:2010, maintenance is defined as "combination of all technical, administrative and managerial actions during the life cycle of an item intended to retain it in, or restore it to, a state in which it can perform the required function" [CEN/TC319, 2010]. There were basically two types of maintenance strategies: corrective maintenance and preventive maintenance. They can also be further divided into sub-categories.

In EN 13306:2010, predictive maintenance is defined as "condition based maintenance carried out following a forecast derived from repeated analysis or known characteristics and evaluation of the significant parameters of the degradation of the item". In contrast to traditional condition based maintenance actions that are based on the available information of the current condition (called "condition based, non-predictive"), the most important aspect of predictive maintenance is the use of methods and models for making a forecast for further condition development and remaining useful life. This means that traditional condition based maintenance recommends maintenance actions based on the information collected through condition monitoring and the focus is on the current condition, compared to establishing a forecast when using predictive maintenance.

In addition, according to EN 13306:2010, preventive maintenance is defined as "Maintenance carried out at predetermined intervals or according to prescribed criteria and intended to reduce the probability of failure or the degradation of the functioning of an item". However, with the development of predictive maintenance, and impact of industry 4.0, it tends to the philosophy “execute at the right time”. Maintenance shall be performed after analytical models predict that enough indictors are present to accurately predict such things as a component's deterioration or certain failure. Simplified, maintenance take place only when necessary. It utilizes real-time data allowing operation team to prioritize and optimize scheduling, which means predictive maintenance may not follow predetermined intervals or prescribed criteria any longer, but the indictors or failure predicted from analytical models.

Therefore, new classification of maintenance strategy has been presented [K. Wang et al., 2015], as shown in Figure 2.2. The new classification separated predictive maintenance from preventive maintenance. This time corrective maintenance, preventive maintenance, and predictive maintenance, are three parallel types.

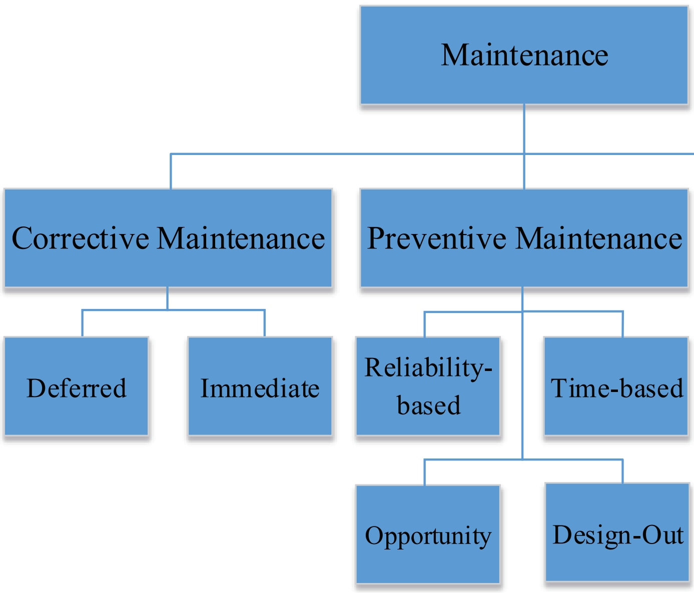

Figure 2.2 Maintenance strategy

2.3.1 Corrective maintenance

Corrective maintenance is a type of maintenance performed to identify and rectify the cause failures for a failed system. It focuses on the identification of failures from the failure phenomenon, which may contain one or more symptom failures [Y. Wang et al., 2014].

Under this strategy, failure is allowed to happen before the maintenance is conducted, which means it is only suitable in the case that the consequence of failure is slight, and some issues like whether the equipment fails, or how long the repair may take does not matter. This type of maintenance can be subdivided into two classes: Deferred and Immediate.

- In deferred corrective maintenance, maintenance is performed in a planned manner, which means correction of fault may not be conducted once the failure occurs but according to the maintenance rules. And this rules may reduce costs or implementation time.
- Immediate corrective maintenance will starts immediately once the failure or degradation is detected.

Corrective maintenance is a failure-driven maintenance, which is undertaken after a breakdown or when obvious failure has been located. The objective of the corrective maintenance is to restore the machine to a state in which it can perform the required function as quickly as possible. Unfortunately, as a primitive maintenance, corrective maintenance do not take account of the loss caused by the unanticipated faults. It is only suited to noncritical areas under the following preconditions: the consequences of failure are slight, there is no safety risk, the failure will be identified quickly, and the repair will be quick. Nevertheless, these preconditions nearly could not be met simultaneously in modern industrial manufacturing fields [C. Fu et al., 2004].

2.3.2 Preventive maintenance

According to European standard EN 13306:2010, preventive maintenance is defined as "Maintenance carried out at predetermined intervals or according to prescribed criteria and intended to reduce the probability of failure or the degradation of the functioning of an item" [CEN/TC319, 2010]. It is carried out at predetermined interval or according to pre-described criteria and intend to reduce the probability of failure or the degradation of a component. This type of maintenance seeks to increase the reliability and availability of equipment through minimizing the number of failures, and avoiding the requirement of unplanned corrective maintenance [De Faria et al., 2015]. In preventive maintenance, equipment can be prevented from faults, since maintenance or correction is performed before failures occur. However, due to preventive action, the replacement of equipment parts may occur prematurely. Nevertheless, preventive maintenance also can not eliminate the occurrence of random catastrophic failure. In addition, with the increase of the quality and complexity of the product, the cost of the frequent maintenance actions is increasingly becoming high. The reliability theory has proven that scheduled maintenance is just suitable for failures that have a clear wear-out characteristic, but unsuitable for random failure [Yu, 2000]. In addition, preventive maintenance raise the number of unnecessary maintenance performance, which may largely increase the cost of maintenance and also cause incidental damage to equipment or components.

Although preventive maintenance may not be the optimum maintenance policy, it still have several advantages as follows [Sullivan et al., 2010]:

- Cost effective in many capital-intensive processes.
- Flexibility allows for the adjustment of maintenance periodicity.
- Increased component life cycle.
- Energy savings.
- Reduced equipment or process failure.
- Estimated 12% to 18% cost savings over reactive maintenance program.

2.3.3 Predictive maintenance

Predictive maintenance includes activities to carry out the appropriate maintenance tasks for maximizing the service life of equipment without increasing the risk of failure taking into account the predicted life of the equipment based on the real states (technical condition) of equipment. By intelligent analysis of (big) data from condition monitoring and operation, predictive maintenance can reduce the costs by reducing the number of unnecessary scheduled time-based maintenance operations.

As an advanced and ideal maintenance policy, predictive maintenance measures parameters in the condition of equipment in order to carry out the appropriate tasks to optimize the service life of machine and processes without increasing the risk of failure. The main function is to collect data with the equipment under operation, discover the information from collected data, and identify potential faults or degradation through historical analysis of similar equipment and knowledge acquired over time. The key is the accuracy of the prediction for the faults or degradation.

Based on the approaches of measuring the symptom of failures, there are two groups of predictive maintenance:

- Statistical-based predictive maintenance.
- Condition-based predictive maintenance.

Statistical-based predictive maintenance is a kind of traditional approach. It is based on the application of statistical or reliability analysis of equipment failure. Under statistical-based predictive maintenance, the objective is to achieve minimum total cost through fixed statistically optimal maintenance intervals to replace or overhaul equipment or components [Mann et al., 1995]. The key of this approach is to apply principles of statistical process to determine when the maintenance shall be performed in the future. Condition-based predictive maintenance involves the application of condition monitoring for the equipment, and predict when, where, which components may have potential failures according to current condition, historical condition, or condition in the future. This type of maintenance mainly depends on continuous or periodic monitoring conditions of equipment to detect the signs of failure and make a maintenance decisions [K. Wang, 2016]. The main challenges of condition-based predictive maintenance in the last a few decades is the access to the necessary data and information for condition monitoring and forecast, and the accuracy of failures or degradation prediction. Fortunately, with the rapid development of condition-monitoring technology and AI techniques in the last few years, the advancement of condition-based predictive maintenance is greatly facilitated. Especially in the industry 4.0 concept, industrial equipments are connected as a community and autonomously exchanging information, which means plentiful industrial data can be obtained conveniently for condition-based maintenance.

Comparing with corrective and preventive maintenance strategies, predictive maintenance has following advantages:

- Equipment that requires maintenance is only shut down before imminent failure.
- Reducing the total time spent maintaining equipment
- Reducing maintenance costs by avoiding catastrophe damage.

- Generating minimal interference in equipment during the operation.
- Increasing availability and reliability of machines.
- Extending life of equipment and processes.

On the down side, the cost of the equipment for condition monitoring needed for predictive maintenance is often high. In addition, the technological capability to accurately interpret condition monitoring data and the ability to acquire and integrate all requisite information is quite hard to achieve previously. In practice, predictive maintenance has a high upfront cost for investment, and management does not readily see savings in the future.

2.4 Predictive maintenance: State-of-the-art

As the most popular and modern maintenance policy, predictive maintenance measures parameters in the condition of equipment carry out the appropriate tasks to optimize the service life of machines and processes without increasing the risk of failure [Garcia et al., 2006]. The methodology is derived from the objective phenomenon: when equipment begins to fail, various types of signs, like fluctuation in temperature, vibration, or noise, can be detected if sharp eyes, ears, and noses are used to sense the failure precursors [Hashemian and Bean, 2011]. Nowadays, in the age of rapid technological advancement, sensors are now advanced enough to play the role of sharp eyes, ears, and noses. Actually, the sensitivity of advanced sensors may be hundreds of times better than human's. However, it does not mean the impetus to predictive maintenance, since we also need to extract information from these signs, discover the knowledge behind the information, and acquire the detail about the potential failures, just like human brain. And how to extract features effectively and appropriately from collected data, identify and classify failures inerrably, and predict the potential faults or degradation precisely for equipment are always hot issues in predictive maintenance.

Many effective and useful methods have been presented, researched and applied for feature extraction or signal analysis in predictive maintenance. These methods can mainly be divided into three types: time domain, frequency domain, and time-frequency domain. The essence of time domain method is an analysis of a waveform. From a mathematical perspective, the waveform of a signal is a chronological sequence of the value of a random variable [Jahnke, 2015]. In frequency domain analysis, signal is analysed with respect to the frequency through certain transformation, which can reveal the distribution of the frequency and filter the noise of signal conveniently. As we discussed above, the objective of predictive maintenance is to evaluate the state or condition of the engineering system through certain signs, to be more specific, the collected signals. The changes of these signals are usually time based issues. However, in many cases, the most distinguished information is hidden in the frequency content of signals, so signal processing techniques in a time-frequency domain is widely used in both diagnosis and prognosis [Jahnke, 2015]. Since feature extraction is a highly subjective problem in nature, the best method normally depends on the practical problem. Table 2.1 lists some common feature extraction methods applied for predictive maintenance based on literature review.

Table 2.1 Application of feature extraction in predictive maintenance

Reference	Main Data Type	Target	Method	Type
[Zarei et al., 2014]	Vibration signal	Bearing fault	Neural Network	Time Domain
[J. Yang et al., 2007]	Vibration signal	Rolling elements	Fractal Dimension	Time Domain
[Abdennadher et al., 2010]	Electrical signal	Electrolytic capacitor	Z Transformation	Frequency Domain
[Liu et al., 2010]	Energy signal	Induction motors	Fast Fourier Transform	Frequency Domain
[Taj et al., 2017]	Maintenance data	Subsystem of a cable plant	Laplace Transformation	Frequency Domain
[Mehta et al., 2015]	Vibration signal	Machine tool's spindle	Short Time Fourier Transform	Time-Frequency Domain
[Z. Zhang et al., 2013]	Vibration signal	Blower	Wavelet Transformation	Time-Frequency Domain
[Wu et al., 2012]	Vibration signal	Gear faults	Hilbert Huang Transformation	Time-Frequency Domain
[C. Wang et al., 2008]	Vibration acceleration signal	Diesel valve trains	Wigner Ville Distribution	Time-Frequency Domain

After processing signals and extracting the features, the more crucial and challenging thing is how to identify current failures and predict potential faults. Actually, this is the key to predictive maintenance. For decades, how to construct an analytical model that can map extracted features or collected data to the indictors or information, which can accurately predict component's deterioration or certain failure, requires elaborate engineering. In recent years, with the dramatic development of AI techniques, they have been increasingly applied to machine diagnosis or prognosis, and have shown improved performance over conventional approaches [Yan, 2014]. In literature review, artificial neural networks (ANNs) are one of the most popular AI techniques for machine diagnosis and prognosis.

ANNs are computational models that mimic the structure of human brains. The purpose is to make machine learn from data and think like human being. An artificial neural network consists of simple processing units as neurons connected in a complex layer structure. This structure enables the model to approximate a complex non-linear function with types of inputs and outputs. Through adjusting the weights of the processing units, ANNs have the ability to learn unknown functions [Nourmohammadzadeh and Hartmann, 2015]. Various types of ANNs were applied and researched for faults diagnosis or prognosis, such as back propagation neural network (BPNN) [Jafar et al., 2010; Rohani et al., 2011], radial basis function neural networks (RBFNN) [G. Xiong et al., 2013], counter propagation neural networks (CPNN) [Phillips et al., 2015], cascade correlation neural network (CCNN) [Phillips et al., 2015], LVQ [C.-C. Wang and James Too, 2002], self-organising map (SOM) [Rai and Upadhyay, 2017; C.-C. Wang and James Too, 2002], recurrent neural networks (RNN) [Yam et al., 2001], dynamic wavelet neural network (DWNN) [Vachtsevanos and Wang, 2001], polynomial neural network (PNN) [Zou et al., 2017], neuro-fuzzy network (NFN) [W. Wang, 2007],and dynamic neural network [Abed et al., 2014]. Another type of popular AI techniques for fault diagnosis and prognosis is expert system (ES). Different from ANNs, which mainly learn and discover knowledge through training on observed data with inputs and outputs, ES utilize domain expert knowledge in a computer program with an automated inference engine to perform reasoning for problem solving. Based the review made by [Jardine et al., 2006], according to the main reasoning, the expert systems applied in the area of machinery diagnostics can be divided into three groups rule-based reasoning [Yoon et al., 1992], case-based reasoning [Ziyan et al., 2003] and model-based reasoning [Baig and Sayeed, 1998]. Expert system is suitable for problems that usually solved by human specialists. However, it is difficult to obtain domain knowledge and convert it to rules [David and Krivine, 1987], and once built, an expert system do not have the ability to handle new situations that may not be covered explicitly in the knowledge bases. Furthermore, when the number of rules increases dramatically, the situation may cause the "combinatorial explosion", which is involved in computation problems [Y. Peng et al., 2010]. Other models like support vector machines (SVM) [Konar and Chattopadhyay, 2011; Rai and Upadhyay, 2017], fuzzy logic system (FLS) [Sobanski, 2014], decision tree (DT) [Muralidharan and Sugumaran, 2013] , and random forest [B.-S. Yang, Di, et al., 2008] are also well-know and useful in certain fields. Table 2.2 lists all the literature reviewed during the research, which includes most of the important and popular AI techniques successfully and widely applied in fault diagnosis and prognosis in recent years.

Table 2.2 Application of AI techniques in predictive maintenance

Reference	**Target**	**Method**
[Kim et al., 2012]	Prognosis of bearing faults	SVM
[Konar and Chattopadhyay, 2011]	Bearing fault detection in induction motor	SVM
[Yoon et al., 1992]	Fault diagnostics of crude units	Rule-based ES
[Baig and Sayeed, 1998]	Fault diagnosis of twin-spool turbofans	Model-based ES
[Ziyan et al., 2003]	Vehicle fault diagnostics	Case-based ES
[B.-S. Yang, Di, et al., 2008]	Diagnosis of induction motors	Random forest
[Muralidharan and Sugumaran, 2013]	Diagnosis of mono-block centrifugal pump	DT
[Sobanski, 2014]	Diagnosis of voltage inverter	FLS
[B.-S. Yang, Oh, et al., 2008]	Predicting the operating conditions of machine	DT & Neuro-fuzzy System

[Ghate and Dudul, 2011]	Fault classification for three-phase induction motor	CCNN
[G. Xiong et al., 2013]	Fault diagnosis of large-scale power systems	RBFNN
[D. Peng et al., 2014]	Diagnosis of turbine generator unit	CPNN
[Deuszkiewicz and Radkowski, 2003]	Condition assessment of power transmission units	CPNN
[Phillips et al., 2015]	Diagnosis mining trucks	CCNN
[Abed et al., 2014]	Diagnosis of brushless DC motor	Dynamic Neural Network
[C.-C. Wang and James Too, 2002]	Rotating machine fault detection	SOM & LVQ
[Rai and Upadhyay, 2017]	Degradation assessment of bearing	SOM & SVM
[Yam et al., 2001]	Equipment deterioration detection in power plants	RNN
[Vachtsevanos and Wang, 2001]	Prognosis of bearing failures	DWNN
[Zou et al., 2017]	Diagnosis of a transformer	PNN
[Unal et al., 2014]	Defects identification for rolling bearings	GA-ANN
[W. Wang, 2007]	Prediction of spur gear condition value one step ahead	NFN
[D.-M. Yang et al., 2002]	Diagnosis of motor bearing	BPNN
[Khomfoi and Tolbert, 2007]	Fault diagnosis for multilevel inverter drive	BPNN
[J.-l. Zhao and Zhao]	Fault diagnosis for missile electronic command system	BPNN
[Rohani et al., 2011]	Predicting repair and maintenance cost	BPNN
[Huijie et al., 2015]	Fault diagnosis in digital circuits	BPNN
[Saravanan and Ramachandran, 2010]	Gear box fault diagnosis	BPNN
[Mehrjoo et al., 2008]	Damage detection of truss bridge joints	BPNN

[Din and Marnerides, 2017]	Power load mapping	DNN
[L. Li and Dai Guilan, 2017]	Fault classification for semiconductor manufacturing process	DNN
[L. Wang et al., 2017]	Identification for failures in wind turbine gearbox	DNN
[Jia et al., 2016]	Processing massive fault data to evaluate the health condition rotating machinery	SAE
[Galloway et al., 2016]	Fault classification for turbine's generator by mining information from spectrograms	SAE
[Sun et al., 2016]	Fault diagnosis in induction motor	SAE
[C. Lu et al., 2017]	Fault diagnosis for components in rotary machinery	SAE
[Guo et al., 2016]	Dimension reduction for intelligent bearing condition monitoring	SAE
[F. Zhou et al., 2017]	Fault classification for machinery equipment in multimode	SAE
[Junbo et al., 2015]	Roller bearing fault diagnosis	SAE
[Huijie et al., 2015]	Fault diagnosis of hydraulic pump	SAE
[K. Li and Wang, 2015]	Spacecraft fault diagnosis	SAE
[Gan and Wang, 2016]	Fault location and severity ranking in rolling-element bearing	DBN
[Tamilselvan and Wang, 2013]	Evaluating the health state of aircraft engine and electric power transformer	DBN
[H. Shao et al., 2015]	Identification of rolling bearing faults	DBN
[AlThobiani and Ball, 2014]	Fault diagnosis of the valves in reciprocating compressors	DBN
[Fink and Weidmann, 2013]	Predicting railway operations failures	DBN
[S. Shao et al., 2016]	Fault diagnosis of induction motor	DBN
[Y. Fu et al., 2015]	Cutting states monitoring	DBN
[M. Ma et al., 2016]	Bearing degradation assessment	DBN
[R. Zhao, Wang, et al., 2016]	Prediction of tool wear in a high speed CNC machine	LSTM
[de Bruin et al., 2017]	Diagnosis time for railway track circuit	LSTM

[Liao and Ahn, 2016]	Asset health assessment	LSTM
[Yuan et al., 2016]	Fault diagnosis and remaining useful life estimation for aero engine	LSTM
[Malhotra et al., 2016]	Prognosis for a pulveriser mill from multi-sensor time-series data	LSTM
[Aydin and Guldamlasioglu, 2017]	Engine condition monitoring	LSTM
[ElSaid et al., 2016]	Predicting excess vibration events in aircraft engines	LSTM
[Janssens et al., 2016]	Fault detection for rotating machinery	CNN
[Z. Chen et al., 2015]	Diagnosis for a two-stage transmission gearbox	CNN
[Babu et al., 2016]	RUL estimation from multi-variate time series sensor signals	CNN
[L. Zhang et al., 2016]	Road crack detection for transportation maintenance	CNN
[W. Zhang et al., 2018]	Bearing fault diagnosis under noisy environment and different working load	CNN
[Ding and He, 2017]	Spindle bearing fault diagnosis	CNN
[Weimer et al., 2016]	Visual defect detection	CNN

In summary, according to the aforementioned fields of literature, it is found that most conventional machine learning models such as BPNN, SVM, RBFNN, and SOM, have the ability to discover the current faults information and deal with the diagnosis problem when the data is not massive and features can be appropriately extracted. However, when the objective is industrial raw data, big data, and the target is to accurately evaluate the degradation or predict the failures of equipment, one may face challenges or 'bottlenecks' to implement predictive maintenance in those situations. In next section, deep learning-based fault identification and prediction approaches will be introduced along with their superiorities in the implementation of predictive maintenance to solve those 'bottlenecks'.

2.5 Deep learning-based fault identification and prediction

The development of modern industry has caused highly increased complexity in both industrial machinery and production systems, which make it difficult or almost impossible to identify and predict failure conditions in a timely manner with conventional methods [Y.

Peng et al., 2010]. Simultaneously, machine learning techniques have facilitated the advancement in many aspects of modern society, from face recognition [C.-Y. Lu et al., 2013] to web search [P.-S. Huang et al., 2013], from cancer prognosis [Kourou et al., 2015] to AlphaGo [Silver et al., 2016], including the application in machinery fault diagnosis and prognosis. However, it is not easy to apply machine learning techniques in practice for predictive maintenance directly, since conventional machine learning techniques were usually limited by their ability to process natural data in their raw form [LeCun et al., 2015]. In addition, to forecast a potential failure in future is much more challenging than diagnosis due to the absence of data about working condition in future. Therefore, under this background, data-driven model with high complexity may be required to evaluate the development of certain faults or degradations. Of course, it is not so easy to extract such high-level, abstract features from kinds of features in different domains or raw data directly. More recently, as a latest research field of machine learning, deep learning has accelerated its application in fault identification and prediction [Gan and Wang, 2016]. This section was conceived to detail deep learning-based fault identification and prediction approaches along with the superiorities of deep learning in predictive maintenance, which could provide a helpful guidance to select suitable methods during the implementation of predictive maintenance.

2.5.1 Trends in deep learning

As a type of machine learning, an approach of AI, deep learning techniques focus on the construction of deep hierarchical models for machine to learn from data. Since the easiest way to know about deep learning may be the understanding of some historical context, the following trends of deep learning are identified [Goodfellow et al., 2016]:

- Deep learning has had a long and rich history, but has gone by many names reflecting different philosophical viewpoints, such as ‘cybernetics’ in the 1940s to 1960s, and ‘connectionism’ in 1980s to 1990s, and has waxed and waned in popularity.
- Deep learning has become more useful as the amount of available training data has increased.
- Deep learning models have grown in size over time as computer hardware and software infrastructure for deep learning has improved.

- Deep learning has solved increasingly complicated applications with increasing accuracy over time.

AI has been an active field of research and part of our imaginations since computers became available and popular in real life. To be more specific, the term was proposed at the Dartmouth Conferences in 1956 in the United States by John McCarty and colleagues [McCarthy et al., 2006]. The idea is to make machines be able to carry out tasks in a way that we would consider “smart” or intelligent. The development and spread of AI in the last few decades can be found in the number of centres, labs, and graduate programs in universities and companies, as well as types of researches and the growing number of publications in journals, books, and conference proceedings worldwide [Cantu-Ortiz, 2014].

In 1980, the first machine learning workshop was held at Carnegie Mellon University (CMU). This workshop and the subsequent publication in 1983 of the first volume of machine learning gave the field a clear identity and a sense of direction, which in turn stimulated the rapid growth that has continued unabated since then [Michalski and Kodratoff, 1990]. The most basic is to build computer programs to parse data and make them have the ability to acquire their own knowledge or make a determination or prediction about something in the world through input data.

Deep learning hypothesizes that in order to learn high-level representations of data, a hierarchy of intermediate representations are required [Palm, 2012]. It makes deep learning be easily explained in contrast to shallow learning, whose archetypical learning model might be a feedforward neural network with one input, hidden and output layer respectively. Actually, initial deep learning methods grow from this kind of shallow learning model, artificial neural network with back-propagation (BP) learning method. However, researchers started to realize that it was extremely difficult to apply BP-based training method for deep neural networks with multiple hidden layers in practice by the late 1980s. In 1991, a research made by Hochreiter, which has been considered as a milestone of explicit deep learning today, formally identified the major cause of the difficulty: Typical deep neural networks will suffer from the problem of gradients vanishing or exploding. With standard activation functions, cumulative error signals from back-propagation either decay or explode exponentially through the number of layers [Hochreiter, 1991]. Since then, research about deep learning was motivated by this insight. Over the years, several popular approaches, which can partially overcome this fundamental issue, were developed [Schmidhuber, 2015]:

- Alleviating the problem through unsupervised pre-training for a hierarchy network. The basic idea is to train each layer in unsupervised fashion to predict its next input. This greatly facilitates subsequent supervised credit assignment through BP. One typical type of deep learning approaches is DBN, which is a stack of RBM.
- Long short term memory (LSTM) like networks alleviate the problem through a special architecture unaffected by it. In a LSTM network employed Constant Error Carousels as activation and identity functions to discover the importance of events that happened thousands of discrete time steps ago.
- The hardware advances make today's computers, especially GPU-based computers, have a million times the computational power of the early 1990s. This allows for propagating errors a few layers further down within reasonable time, which means the standard BP training is feasible a few layers deeper than when the gradients vanishing problem was recognized.
- The space of neural network weight matrices can also be searched without relying on error gradients, thus avoiding the problem of gradients vanishing altogether. Actually, random weight guessing sometimes works better than more sophisticated methods [Hochreiter and Schmidhuber, 1996].

Those approaches facilitate the development of deep learning and eventually formulate types of deep learning architectures, which are widely leveraged in fault diagnosis and prognosis today. In the following sections, the application of those deep learning approaches will be introduced along with their superiorities in certain cases or with prerequisites.

2.5.2 Application of deep learning in predictive maintenance

In practice, it is not easy to apply AI techniques due to the lack of efficient procedures to obtain training data and specific knowledge, which are required to train the models [Jardine et al., 2006]. As reported by LeCun [2015], conventional machine-learning techniques were usually limited by their ability to process natural data in their raw form. For decades, to construct a machine learning system required elaborate engineering and considerable domain expertise to design a feature extraction and selection system that transformed the raw data into a suitable internal representation or feature vector from which the learning subsystem, often a classifier, could detect or classify patterns in the input. More recently, deep learning, as a latest research area of machine learning, has accelerated its application in fault diagnosis and prognosis [Gan and Wang, 2016]. As a branch of machine learning, deep learning is a series of algorithms, which can be applied to model, approximate, or map high-level abstractions in data. The essence of deep learning is about how to compute hierarchical features or representations from the objective data. The family of deep learning approaches have been growing increasingly richer, encompassing variety of neural networks with multiple processing layers, hierarchical probabilistic models, and kinds of unsupervised or supervised feature learning algorithms [Deng and Yu, 2014]. Schmidhuber [2015] summarized all relevant work about deep learning in neural networks and distinguished the shallow and deep learning methods by the depth of their credit assignment paths. The key advantage of deep learning is that the features are not designed by human engineers but learned from data itself through a generalized self-learning procedure. Various deep learning algorithms have been applied successfully in the field of computer vision [Krizhevsky et al., 2012; Ribeiro et al., 2011; Tompson et al., 2014; S. Zhou et al., 2010], speech recognition [G. Hinton et al., 2012; Sainath et al., 2013; Seltzer et al., 2013], and proved the good performance in predicting the activity of potential drug molecules [Hermann et al., 2015], analysing particle accelerator data [Ciodaro et al., 2012], reconstructing brain circuits [Helmstaedter et al., 2013], and predicting the effects of mutations in non-coding DNA on gene expression and disease [H. Y. Xiong et al., 2015]. Currently, various deep learning algorithms, such as deep belief networks (DBN) [Gan and Wang, 2016; Tamilselvan and Wang, 2013] and deep neural networks [L. Wang et al., 2017], have been applied successfully in predictive maintenance. Here, part of relevant research in this area is listed.

Tamilselvan and Wang [2013] presented a multi-sensor health diagnosis method based on DBN. It employs a hierarchical structure with multiple stacked restricted Boltzmann machines (RBM) and works through a layer by layer successive learning process. The method is successfully applied to assess the degradation of aircraft engines and electric power transformers. By comparison with other classification algorithms, such as SVM, SOM, and BPNN, DBN proves a better diagnosis performance for health diagnosis of complex systems.

Wang et al [2017] proposed a method driven by deep neural network to evaluate the conditions of wind turbine gearboxes and identify their impending failures. In the experiment, parallelized stochastic gradient descent is used to accelerate the training process of deep neural networks. To prevent the overfitting of the deep neural networks model, a dropout algorithm is also applied into the deep neural networks training process. According to the comparison with other shallow networks, deep neural networks has the best performance in failures prediction.

Gan and Wang [2016] have presented a hierarchical diagnosis network by collecting RBM for fault pattern recognition in rolling element bearings. In the hierarchical network, two decision layers are designed to identify fault types and evaluate the degradation respectively. To confirm the effectiveness of the deep neural network, SVM and BPNN were also employed to present a comprehensive comparison. The experiment results showed that DBN is highly reliable for precise multi–stage diagnosis and can overcome the overlapping problem caused by noise and other disturbances.

Jia et al. [2016] proposed an intelligent method based on deep neural network to process the massive fault data and automatically provide accurate diagnosis results for rotating machinery. The authors found that, deep neural networks with deep architectures could be established to mine the useful information from raw data and approximate complex non-linear functions. The effectiveness of the proposed method is validated using datasets from rolling element bearings and planetary gearboxes. These datasets contain massive measured signals involving different health conditions under various operating conditions. The diagnosis results show that the deep neural network has the ability to mine available fault characteristics from the raw data, and obtain superior diagnosis accuracy compared with conventional methods.

2.5.3 Superiority of deep learning for fault identification and prediction

In order to overcome the fundamental problem of gradients vanishing, various types of deep learning algorithms are proposed and developed in the last few decades. As reported by Deng [2012], deep learning today refers to a rather wide class of machine learning techniques and architectures, with the hallmark of employing many layers of non-linear information process stages, which are highly hierarchical in nature, to obtain a better accuracy, performance or deal with issues with high complexity. In this chapter, five deep learning architectures including deep neural network with back-propagation (DNN), stacked autoencoders (SAE), DBN, LSTM, and convolutional neural networks (CNN), which might ideally get rid of some 'bottlenecks' that conventional methods faced during the implementation of predictive maintenance, will be introduced and researched both in theory and practical applications based on literature review to illuminate the superiorities of deep learning approaches in certain issues for predictive maintenance. It could offer a guidance to select suitable deep learning models for practical applications. Table 2.2 listed all practical applications of deep learning methods for fault diagnosis and prognosis along with their targets in literature. Table 2.3 listed the superiority of five types of deep learning algorithms for predictive maintenance.

Table 2.3 Superiority of deep learning for predictive maintenance

Architecture	Superiority for predictive maintenance
DNN	Degradation mapping, and failures identification, when enough history data could be obtained, and the complexity of target issue is relatively high.
SAE	Fault characteristics mining, extracting features or hidden information about failures from the raw input data and subsequently dividing them into different levels, dimensionality reduction, and discovering discriminative information about failures when the input dimensionality is large.
DBN	Energy-based models enable DBN to mine information hidden behind highly coupled inputs, which makes DBN a feasible method for fault diagnosis and prognosis when the target condition is beyond the historical data. In addition, it also has the ability to discover the discriminative information about failures when the input dimensionality is large.
LSTM	By stacking memory cells, information of previous inputs can be kept in the output to some degree, carried by cell state, which makes LSTM an outstanding tool to mimic time series.
CNN	Strong capacity to discover knowledge behind large data especially

	for image-based data. And due to the ability to learn complex and robust representation via its convolutional layer, filters in convolutional layers may extract local patterns in raw data and further build complex patterns for machine health monitoring through stacking these convolutional layers.

As an evolutionary version of BPNN, DNN is also widely applied in fault diagnosis and prognosis to pursue higher accuracy for faults detection, classification, or prediction. Compared with other conventional machine learning algorithms, the superiority of DNN is in degradation mapping, and failures identification, when enough history data could be obtained, and the complexity of target issue is relatively high. Evidence can also be found in practical applications [Din and Marnerides, 2017; L. Li and Dai Guilan, 2017; L. Wang et al., 2017]. For example, in [L. Wang et al., 2017], the authors employed DNN to map the lubricant pressure for wind turbine gearbox and subsequently identify the impending failures. They also offers evidence that DNN can provide better performance in prediction compared with other five data-mining algorithms, K-nearest neighbours, Least Absolute Shrinkage and selection operator, Ridge Regression, SVM, and BPNN. In [L. Li and Dai Guilan, 2017], a DNN with multiple hidden layers is applied for fault classification in a semi-conductor manufacturing process. The result in that paper also shows that DNN is more competitive in the aspect of convergence speed and outperforms other conventional approaches, such as multilayer perceptron, SVM, and logistic regression.

SAE has been widely and successfully applied to dimensionality reduction in many research fields [Shin et al., 2013; Zabalza et al., 2016]. Due to the clear hierarchical relationship between each two layers, SAE has the capacity to implement fault characteristics mining, extract features or hidden information about failures from the raw input data, and divide them into different levels. Actually, dimensionality reduction was one of the most original applications of deep learning, which was also one of the early motivations for developing autoencoders [G. E. Hinton and Salakhutdinov, 2006]. Models with smaller spaces could consume less memory, runtime and computation load for the system. In general, for predictive maintenance, SAE has the superiority to capture the main variations and discover the discriminative information about failures when the input dimensionality is large. Some practical applications based on literature review can also support the theoretical speculation. In [Jia et al., 2016], SAE was widely applied to process the massive fault data and evaluate the health condition of the rolling element bearings and planetary gearboxes in rotating machinery. Galloway [2016] trained a deep neural network of SAE with spectrograms constructed from raw vibration data directly instead of feature extraction. The result shows that SAE can learn the response of the tidal turbine under variable loading conditions and identify faults within the turbine's generator. According to the comparison with other feature-based method, such as SVM, decision tree and KNN classifiers, which are trained after extracting features from vibration data, SAE has even better performance for faults classification in that case.

DBN, as a type of energy-based models, is constructed through training and stacking several layers of RBM, which makes the learning process correspond to modifying energy function so that its shape has desirable properties [Bengio, 2009]. Literatures indicate these layer-by-layer nonlinear learning networks with fine-tuning procedure could enable DBN to capture intrinsic characteristics about potential failures from the massive data. In [Gan and Wang, 2016], the authors applied two layers of DBN to deal with the non-stationary property of vibration signals for fault location and severity ranking in rolling-element bearing. The result shows that DBN has the ability to discover the weak links of mechanical system and provide effective information about failures. The authors in that literature also provide a comprehensive evidence for the accuracy and efficiency of DBN by utilizing BPNN and SVM for comparison. The comparison results demonstrate that DBN has a better performance than BPNN and SVM, especially in fault location and classification. Tamilselvan and Wang [2013] applied DBN to evaluate the health state of aircraft engine and electric power transformer, respectively. In the experiment, the diagnosis performances of the DBN were compared with SVM, BPNN, and SOM. Case study results indicated that DBN generally results in a better diagnosis performance for health diagnosis in complex systems, compared to other classification methods. Besides, DBN has proved its outstanding performance in many fields such as prediction of chaotic time series [Kuremoto et al., 2014], traffic flow prediction [W. Huang et al., 2014], financial business prediction [Ribeiro and Lopes, 2011], short-term prediction of drought [J. Chen et al., 2012], and retrieval term prediction [Q. Ma et al., 2014]. For example, Kuremoto et al. [2014] successfully employed DBN to predict the Lorenz chaos well-known with its "butterfly effect", which can indicate the sensitive dependence on initial conditions of chaos. In [Q. Ma et al., 2014], DBN was applied to predict retrieval terms from relevant and surrounding words, or descriptive texts. To determine the effectiveness of DBN, the authors tested it along with baseline methods such as multi-layer perceptron and SVM for comparison. The experimental results showed that DBN has far higher prediction precisions than the others. In Chapter 5, a novel application of DBN to predict backlash error in a machining center when the target condition is beyond historical data will be introduced in detail. That case study could also prove the theoretical speculation. In general, due to the pre-training process through unsurprised learning, DBN has the superiority to discover discriminative information about failures when the input dimensionality is large.

As a type of RNN, LSTM is constructed through stacking memory cells, which can keep information of previous inputs in the output to some degree. For this reason, LSTM is an outstanding tool to mimic time series and has been successfully applied in various applications, such as speech recognition [Graves et al., 2013], information retrieval [Palangi et al., 2016], protein disorder prediction [Hanson et al., 2016], handwriting recognition, and processing acoustic sequences [Sak and Senior, 2017]. As reported by Zhao et al. [2016], many machinery data is obtained from sensor data, which is highly in nature time series. Therefore, in predictive maintenance, LSTM is also a popular and useful model to discover temporal information from sequential data, especially when the issue is highly related with time series. For example, Zhao et al. [2016] leveraged LSTM to mine the information hidden in raw sensory data and successfully predicted the corresponding tool wear for a high speed CNC machine. The authors also compared LSTM with other benchmark methods such as linear regression, support vector machine regression (SVMR), and multi-layer perception (MLP). The results shows that LSTM has the ability to learn meaningful representations from raw signal and better performance than conventional methods in the issues with high temporal dependency. In [de Bruin et al., 2017], LSTM has been applied to identify the fault types and to determine the development of the fault severity over time for railway track circuit. Malhotra [2016] proposed a LSTM-based encoder-decoder model to estimate the remaining useful life of a pulveriser mill from multi-sensor time-series data. The experiment results also show the outstanding performance of LSTM to reconstruct the time-series corresponding to healthy state.

CNN is a kind of hierarchical multi-layered model with a very strong capacity to discover knowledge behind large data especially for image-based data since vision is highly hierarchically organised. For predictive maintenance, in some scenarios, the information or signs about failures can also be perceived from data in 2D format, such as time-frequency spectrum and pictures. As reported by Zhao et al. [2016], due to the ability to learn complex and robust representation via its convolutional layer, filters in convolutional layers may extract local patterns in raw data and further build complex patterns for machine health monitoring through stacking these convolutional layers, which makes CNN an ideal tool when the target is image-based data.

In Chapter 5, a novel DBN-based error prediction approach will be introduced in a case study for backlash error prediction in a machining center, which could prove the superiority of DBN in the situation when target condition is beyond the historical data. In that case, a comprehensive comparison will also be provided to validate the effectiveness of DBN. The advantages of DNN, SAE, and LSTM will also be discussed during the experiment in Chapter 6, in which the effectiveness of DNN-based degradation assessment, SAE-based feature representation, and LSTM-based anomaly identification with time series will also be validated.

2.6 Summary

This chapter has introduced the development of predictive maintenance, along with various techniques and models that could be chosen to implement it. The impact of Industry 4.0, and trend to leverage deep learning approaches are also explained to the readers for a general understanding of the background.

In sense, deep learning could be a feasible and effective method to solve certain issues or challenges during the implementation of predictive maintenance, according to its application in other field such as computer vision [Krizhevsky et al., 2012; Ribeiro et al., 2011; Tompson et al., 2014; S. Zhou et al., 2010], speech recognition [G. Hinton et al., 2012; Sainath et al., 2013; Seltzer et al., 2013], and biological science [H. Y. Xiong et al., 2015]. Therefore, it is important and imperative to establish a framework for predictive maintenance concerning the Industry 4.0 concept and deep learning approaches, which can offer an overall understanding and effective guidance for researchers and practitioners to implement predictive maintenance in the new situation. This chapter also provides a systematic review about deep learning approaches applied in fault identification and prediction. Five types of deep learning architectures including DNN, SAE, DBN, LSTM and CNN, which can improve or solve some 'bottlenecks' in predictive maintenance are discussed. The chapter also provides the superiorities of these architectures in certain issues for predictive maintenance, which could offer an effective guidance for researchers and practitioners to select appropriate deep learning architectures during the implementation of predictive maintenance in Industry 4.0 era.

References

Abdennadher, K., Venet, P., Rojat, G., Rétif, J.-M., and Rosset, C. (2010). A real-time predictive-maintenance system of aluminum electrolytic capacitors used in uninterrupted power supplies. *IEEE Transactions on Industry Applications, 46*(4), 1644-1652.

Abed, W., Sharma, S., and Sutton, R. (2014). *Diagnosis of bearing fault of brushless DC motor based on dynamic neural network and orthogonal fuzzy neighborhood discriminant analysis.* 2014 UKACC International Conference on the Control (CONTROL).

AlThobiani, F., and Ball, A. (2014). An approach to fault diagnosis of reciprocating compressor valves using Teager–Kaiser energy operator and deep belief networks. *Expert systems with applications, 41*(9), 4113-4122.

Arslan, A. K., Colak, C., and Sarihan, M. E. (2016). Different medical data mining approaches based prediction of ischemic stroke. *Computer Methods and Programs in Biomedicine, 130*, 87-92.

Aydin, O., and Guldamlasioglu, S. (2017). *Using LSTM networks to predict engine condition on large scale data processing framework.* 2017 4th International Conference on Electrical and Electronic Engineering (ICEEE).

Babu, G. S., Zhao, P., and Li, X.-L. (2016). *Deep convolutional neural network based regression approach for estimation of remaining useful life.* Paper presented at the International conference on database systems for advanced applications.

Baheti, R., and Gill, H. (2011). Cyber-physical systems. *The impact of control technology, 12*, 161-166.

Baig, M. F., and Sayeed, N. (1998). Model-based reasoning for fault diagnosis of twin-spool turbofans. *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering, 212*(2), 109-116.

Bengio, Y. (2009). Learning deep architectures for AI. *Foundations and trends® in Machine Learning, 2*(1), 1-127.

Cantu-Ortiz, F. J. (2014). Advancing artificial intelligence research and dissemination through conference series: Benchmark, scientific impact and the MICAI experience. *Expert systems with applications, 41*(3), 781-785.

CEN/TC319. (2010). EN 13306:2010. Maintenance—maintenance terminology: European Standard.

Chaves, L. W., and Nochta, Z. (2011). Breakthrough towards the internet of things *Unique Radio Innovation for the 21st Century* (pp. 25-38): Springer.

Chen, J., Jin, Q., and Chao, J. (2012). Design of deep belief networks for short-term prediction of drought index using data in the Huaihe river basin. *Mathematical Problems in Engineering, 2012*.

Chen, Z., Li, C., and Sanchez, R.-V. (2015). Gearbox fault identification and classification with convolutional neural networks. *Shock and Vibration, 2015*.

Ciodaro, T., Deva, D., De Seixas, J., and Damazio, D. (2012). *Online particle detection with neural networks based on topological calorimetry information.* Paper presented at the Journal of Physics: Conference Series.

David, J.-M., and Krivine, J.-P. (1987). Three artificial intelligence issues in fault diagnosis: declarative programming, expert systems, and model-based reasoning. *Fault detection and reliability: knowledge based and other approaches*, 19-27.

de Bruin, T., Verbert, K., and Babuška, R. (2017). Railway track circuit fault diagnosis using recurrent neural networks. *IEEE transactions on neural networks and learning systems, 28*(3), 523-533.

De Faria, H., Costa, J. G. S., and Olivas, J. L. M. (2015). A review of monitoring methods for predictive maintenance of electric power transformers based on dissolved gas analysis. *Renewable and sustainable energy reviews, 46*, 201-209.

Deng, L. (2012). Three classes of deep learning architectures and their applications: a tutorial survey. *APSIPA transactions on signal and information processing*.

Deng, L., and Yu, D. (2014). Deep Learning: Methods and Applications. *Foundations and Trends® in Signal Processing, 7*(3–4), 197-387. doi:10.1561/2000000039

Deuszkiewicz, P., and Radkowski, S. (2003). On-line condition monitoring of a power transmission unit of a rail vehicle. *Mechanical Systems and Signal Processing, 17*(6), 1321-1334.

Din, G. M. U., and Marnerides, A. K. (2017). *Short term power load forecasting using Deep Neural Networks.* Networking and Communications (ICNC), 2017 International Conference on the Computing.

Ding, X., and He, Q. (2017). Energy-fluctuated multiscale feature learning with deep convnet for intelligent spindle bearing fault diagnosis. *IEEE Transactions on Instrumentation and measurement*.

Drath, R., and Horch, A. (2014). Industrie 4.0: Hit or hype?[industry forum]. *IEEE industrial electronics magazine, 8*(2), 56-58.

ElSaid, A., Wild, B., Higgins, J., and Desell, T. (2016). *Using LSTM recurrent neural networks to predict excess vibration events in aircraft engines.* 2016 IEEE 12th International Conference on the e-Science (e-Science).

Fink, O., and Weidmann, U. (2013). *Using deep belief networks for predicting railway operations failures.* Paper presented at the Advances in Risk and Reliability Technology Symposium.

Fu, C., Ye, L., Liu, Y., Yu, R., Iung, B., Cheng, Y., and Zeng, Y. (2004). Predictive maintenance in intelligent-control-maintenance-management system for hydroelectric generating unit. *IEEE transactions on energy conversion, 19*(1), 179-186.

Fu, Y., Zhang, Y., Qiao, H., Li, D., Zhou, H., and Leopold, J. (2015). Analysis of feature extracting ability for cutting state monitoring using deep belief networks. *Procedia CIRP, 31*, 29-34.

Galloway, G. S., Catterson, V. M., Fay, T., Robb, A., and Love, C. (2016). Diagnosis of tidal turbine vibration data through deep neural networks. *In: Proceedings of the Third European Conference of the Prognostics and Health Management Society 2016.*(PHM Society), 172-180.

Gan, M., and Wang, C. (2016). Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings. *Mechanical Systems and Signal Processing, 72*, 92-104.

Garcia, M. C., Sanz-Bobi, M. A., and del Pico, J. (2006). SIMAP: Intelligent System for Predictive Maintenance: Application to the health condition monitoring of a windturbine gearbox. *Computers in Industry, 57*(6), 552-568.

Ghate, V. N., and Dudul, S. V. (2011). Cascade neural-network-based fault classifier for three-phase induction motor. *IEEE Transactions on Industrial Electronics, 58*(5), 1555-1563.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*: MIT press.

Gorecky, D., Schmitt, M., Loskyll, M., and Zühlke, D. (2014). *Human-machine-interaction in the industry 4.0 era.* 2014 12th IEEE International Conference on the Industrial Informatics (INDIN).

Graves, A., Mohamed, A.-r., and Hinton, G. (2013). *Speech recognition with deep recurrent neural networks.* 2013 ieee international conference on the Acoustics, speech and signal processing (icassp).

Guo, L., Gao, H., Huang, H., He, X., and Li, S. (2016). Multifeatures fusion and nonlinear dimension reduction for intelligent bearing condition monitoring. *Shock and Vibration, 2016*.

Han, J., Kamber, M., and Pei, J. (2011). *Data mining: concepts and techniques*: Elsevier.

Hanson, J., Yang, Y., Paliwal, K., and Zhou, Y. (2016). Improving protein disorder prediction by deep bidirectional long short-term memory recurrent neural networks. *Bioinformatics, 33*(5), 685-692.

Hashemian, H. M., and Bean, W. C. (2011). State-of-the-art predictive maintenance techniques. *IEEE transactions on instrumentation and measurement, 60*(10), 3480-3492.

Helmstaedter, M., Briggman, K. L., Turaga, S. C., Jain, V., Seung, H. S., and Denk, W. (2013). Connectomic reconstruction of the inner plexiform layer in the mouse retina. *Nature, 500*(7461), 168-174.

Henriquez, P., Alonso, J. B., Ferrer, M. A., and Travieso, C. M. (2014). Review of automatic fault diagnosis systems using audio and vibration signals. *IEEE Transactions on Systems, Man, and Cybernetics: Systems, 44*(5), 642-652.

Hermann, M., Pentek, T., and Otto, B. (2015). Design principles for Industrie 4.0 scenarios: a literature review. *Technische Universität Dortmund, Dortmund*.

Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A.-r., Jaitly, N., . . . Sainath, T. N. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine, 29*(6), 82-97.

Hinton, G. E., and Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science, 313*(5786), 504-507.

Hochreiter, S. (1991). Untersuchungen zu dynamischen neuronalen Netzen. *Diploma, Technische Universität München, 91*.

Hochreiter, S., and Schmidhuber, J. (1996). Bridging long time lags by weight guessing and “Long Short-Term Memory”. *Spatiotemporal models in biological and artificial systems, 37*, 65-72.

Hochschild, A. (2015). EMPIRE OF COTTON A Global History: NEW YORK TIMES 620 8TH AVE, NEW YORK, NY 10018 USA.

Huang, P.-S., He, X., Gao, J., Deng, L., Acero, A., and Heck, L. (2013). *Learning deep structured semantic models for web search using clickthrough data.* Paper presented at the Proceedings of the 22nd ACM international conference on Conference on information & knowledge management.

Huang, W., Song, G., Hong, H., and Xie, K. (2014). Deep architecture for traffic flow prediction: deep belief networks with multitask learning. *IEEE Transactions on Intelligent Transportation Systems, 15*(5), 2191-2201.

Huijie, Z., Ting, R., Xinqing, W., You, Z., and Husheng, F. (2015). *Fault diagnosis of* *hydraulic pump based on stacked autoencoders.* 2015 12th IEEE International Conference on the Electronic Measurement & Instruments (ICEMI).

Jafar, R., Shahrour, I., and Juran, I. (2010). Application of Artificial Neural Networks (ANN) to model the failure of urban water mains. *Mathematical and Computer Modelling, 51*(9), 1170-1180.

Jahnke, P. (2015). Machine learning approaches for failure type detection and predictive maintenance. *Technische Universität Darmstadt, 19*.

Janssens, O., Slavkovikj, V., Vervisch, B., Stockman, K., Loccufier, M., Verstockt, S., . . . Van Hoecke, S. (2016). Convolutional neural network based fault detection for rotating machinery. *Journal of Sound and Vibration, 377*, 331-345.

Jardine, A. K., Lin, D., and Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing, 20*(7), 1483-1510.

Jia, F., Lei, Y., Lin, J., Zhou, X., and Lu, N. (2016). Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. *Mechanical Systems and Signal Processing, 72*, 303-315.

Junbo, T., Weining, L., Juneng, A., and Xueqian, W. (2015). *Fault diagnosis method study in roller bearing based on wavelet transform and stacked auto-encoder.* Paper presented at the Control and Decision Conference (CCDC), 2015 27th Chinese.

Khomfoi, S., and Tolbert, L. M. (2007). Fault diagnosis and reconfiguration for multilevel inverter drive using AI-based techniques. *IEEE Transactions on Industrial Electronics, 54*(6), 2954-2968.

Kim, H.-E., Tan, A. C., Mathew, J., and Choi, B.-K. (2012). Bearing fault prognosis based on health state probability estimation. *Expert systems with applications, 39*(5), 5200-5213.

Konar, P., and Chattopadhyay, P. (2011). Bearing fault detection of induction motor using wavelet and Support Vector Machines (SVMs). *Applied Soft Computing, 11*(6), 4203-4211.

Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., and Fotiadis, D. I. (2015).

Machine learning applications in cancer prognosis and prediction. *Computational and structural biotechnology journal, 13*, 8-17.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). *Imagenet classification with deep convolutional neural networks.* Paper presented at the Advances in neural information processing systems.

Kuremoto, T., Kimura, S., Kobayashi, K., and Obayashi, M. (2014). Time series forecasting using a deep belief network with restricted Boltzmann machines. *Neurocomputing, 137*, 47-56.

Lasi, H., Fettke, P., Kemper, H.-G., Feld, T., and Hoffmann, M. (2014). Industry 4.0. *Business & Information Systems Engineering, 6*(4), 239.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436-444.

Lee, J., Kao, H.-A., and Yang, S. (2014). Service innovation and smart analytics for industry 4.0 and big data environment. *Procedia CIRP, 16*, 3-8.

Li, K., and Wang, Q. (2015). *Study on signal recognition and diagnosis for spacecraft based on deep learning method.* Paper presented at the Prognostics and System Health Management Conference (PHM), 2015.

Li, L., and Dai Guilan, Z. Y. (2017). A Membership-based Multi-dimension Hierarchical Deep Neural Network Approach for Fault Diagnosis.

Liao, L., and Ahn, H.-i. (2016). Combining Deep Learning and Survival Analysis for Asset Health Management. *International journal of prognostics and health management*.

Liu, Y., Guo, L., Wang, Q., An, G., Guo, M., and Lian, H. (2010). Application to induction motor faults diagnosis of the amplitude recovery method combined with FFT. *Mechanical Systems and Signal Processing, 24*(8), 2961-2971.

Lu, C.-Y., Min, H., Gui, J., Zhu, L., and Lei, Y.-K. (2013). Face recognition via weighted sparse representation. *Journal of Visual Communication and Image Representation, 24*(2), 111-116.

Lu, C., Wang, Z.-Y., Qin, W.-L., and Ma, J. (2017). Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification. *Signal Processing, 130*, 377-388.

Ma, M., Chen, X., Wang, S., Liu, Y., and Li, W. (2016). *Bearing degradation assessment based on weibull distribution and deep belief network.* Paper presented at the International Symposium on Flexible Automation (ISFA).

Ma, Q., Tanigawa, I., and Murata, M. (2014). *Retrieval Term Prediction Using Deep Belief Networks.* Paper presented at the PACLIC.

MacDougall, W. (2014). *Industrie 4.0: Smart manufacturing for the future*: Germany Trade & Invest.

Malhotra, P., TV, V., Ramakrishnan, A., Anand, G., Vig, L., Agarwal, P., and Shroff, G. (2016). Multi-sensor prognostics using an unsupervised health index based on lstm encoder-decoder. *arXiv preprint arXiv:1608.06154*.

Mann, L., Saxena, A., and Knapp, G. M. (1995). Statistical-based or condition-based preventive maintenance? *Journal of Quality in Maintenance Engineering, 1*(1), 46-59.

McCarthy, J., Minsky, M. L., Rochester, N., and Shannon, C. E. (2006). A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955. *AI magazine, 27*(4), 12.

Mehrjoo, M., Khaji, N., Moharrami, H., and Bahreininejad, A. (2008). Damage detection of truss bridge joints using Artificial Neural Networks. *Expert systems with applications, 35*(3), 1122-1131.

Mehta, P., Werner, A., and Mears, L. (2015). Condition based maintenance-systems integration and intelligence using Bayesian classification and sensor fusion. *Journal of Intelligent Manufacturing, 26*(2), 331-346.

Michalski, R., and Kodratoff, Y. (1990). Research in machine learning. *Machine Learning III. An Artificial Intelligence Approach*, 3-30.

Muralidharan, V., and Sugumaran, V. (2013). Feature extraction using wavelets and classification through decision tree algorithm for fault diagnosis of mono-block centrifugal pump. *Measurement, 46*(1), 353-359.

Nourmohammadzadeh, A., and Hartmann, S. (2015). *Fault classification of a centrifugal pump in normal and noisy environment with artificial neural network and support vector machine enhanced by a genetic algorithm.* Paper presented at the International Conference on Theory and Practice of Natural Computing.

Palangi, H., Deng, L., Shen, Y., Gao, J., He, X., Chen, J., . . . Ward, R. (2016). Deep sentence embedding using long short-term memory networks: Analysis and application to information retrieval. *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP), 24*(4), 694-707.

Palm, R. B. (2012). Prediction as a candidate for learning deep hierarchical models of data. *Technical University of Denmark, 5*.

Peng, D., Qian, Y., Zhang, H., and Xia, F. (2014). *Research on fault diagnosis of turbine generator unit based on improved CPN neural network.* Paper presented at 2014 IEEE/SICE International Symposium on the System Integration (SII).

Peng, Y., Dong, M., and Zuo, M. J. (2010). Current status of machine prognostics in condition-based maintenance: a review. *The International Journal of Advanced Manufacturing Technology, 50*(1), 297-313. doi:10.1007/s00170-009-2482-0

Phillips, J., Cripps, E., Lau, J. W., and Hodkiewicz, M. (2015). Classifying machinery condition using oil samples and binary logistic regression. *Mechanical Systems and Signal Processing, 60*, 316-325.

Poovendran, R. (2010). Cyber–physical systems: Close encounters between two parallel worlds [point of view]. *Proceedings of the IEEE, 98*(8), 1363-1366.

Rai, A., and Upadhyay, S. H. (2017). Intelligent bearing performance degradation assessment and remaining useful life prediction based on self-organising map and support vector regression. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 0954406217700180.

Ribeiro, B., Gonçalves, I., Santos, S., and Kovacec, A. (2011). *Deep learning networks for off-line handwritten signature recognition.* Paper presented at the Iberoamerican Congress on Pattern Recognition.

Ribeiro, B., and Lopes, N. (2011). *Deep belief networks for financial prediction.* Paper presented at the Neural Information Processing.

Rohani, A., Abbaspour-Fard, M. H., and Abdolahpour, S. (2011). Prediction of tractor repair and maintenance costs using artificial neural network. *Expert systems with applications, 38*(7), 8999-9007.

Sainath, T. N., Mohamed, A.-r., Kingsbury, B., and Ramabhadran, B. (2013). *Deep convolutional neural networks for LVCSR.* Paper presented at the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing.

Sak, H., and Senior, A. W. (2017). Processing acoustic sequences using long short-term memory (lstm) neural networks that include recurrent projection layers: Google Patents.

Saravanan, N., and Ramachandran, K. (2010). Incipient gear box fault diagnosis using discrete wavelet transform (DWT) for feature extraction and classification using artificial neural network (ANN). *Expert systems with applications, 37*(6), 4168-4181.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks, 61*, 85-117.

Schmidt, R., Möhring, M., Härting, R.-C., Reichstein, C., Neumaier, P., and Jozinović, P. (2015). *Industry 4.0-Potentials for Creating Smart Products: Empirical Research Results.* Paper presented at the Business Information Systems.

Seltzer, M. L., Yu, D., and Wang, Y. (2013). *An investigation of deep neural networks for noise robust speech recognition.* Paper presented at the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing.

Shao, H., Jiang, H., Zhang, X., and Niu, M. (2015). Rolling bearing fault diagnosis using an optimization deep belief network. *Measurement Science and Technology, 26*(11), 115002.

Shao, S., Sun, W., Wang, P., Gao, R. X., and Yan, R. (2016). *Learning features from vibration signals for induction motor fault diagnosis.* Paper presented at International Symposium on the Flexible Automation (ISFA).

Shin, H.-C., Orton, M. R., Collins, D. J., Doran, S. J., and Leach, M. O. (2013). Stacked autoencoders for unsupervised feature learning and multiple organ detection in a pilot study using 4D patient data. *IEEE transactions on pattern analysis and machine intelligence, 35*(8), 1930-1943.

Shrouf, F., Ordieres, J., and Miragliotta, G. (2014). *Smart factories in Industry 4.0: A review of the concept and of energy management approached in production based on the Internet of Things paradigm.* Paper presented at 2014 IEEE International Conference on the Industrial Engineering and Engineering Management (IEEM).

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., . . . Lanctot, M. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature, 529*(7587), 484-489.

Sobanski, P. (2014). Fuzzy-logic-based approach to voltage inverter fault diagnosis in induction motor drive. *Przeglad Elektrotechniczny, 90*(6), 149-153.

Sohrabi, M. K., and Akbari, S. (2016). A comprehensive study on the effects of using data mining techniques to predict tie strength. *Computers in Human Behavior, 60*, 534-541.

Sullivan, G., Pugh, R., Melendez, A. P., and Hunt, W. (2010). *Operations & Maintenance Best Practices-A Guide to Achieving Operational Efficiency (Release 3)*. Retrieved from

Sun, W., Shao, S., Zhao, R., Yan, R., Zhang, X., and Chen, X. (2016). A sparse auto-encoder-based deep neural network approach for induction motor faults classification. *Measurement, 89*, 171-178.

Taj, S., Rizwan, S., Alkali, B., Harrison, D., and Taneja, G. (2017). Reliability Analysis of a Single Machine Subsystem of a Cable Plant with Six Maintenance Categories. *International Journal of Applied Engineering Research, 12*(8), 1752-1757.

Tamilselvan, P., and Wang, P. (2013). Failure diagnosis using deep belief learning based health state classification. *Reliability Engineering & System Safety, 115*, 124-135.

Thuemmler, C., and Bai, C. (2017). Health 4.0: Application of Industry 4.0 Design Principles in Future Asthma Management *Health 4.0: How Virtualization and Big Data are Revolutionizing Healthcare* (pp. 23-37): Springer.

Tompson, J. J., Jain, A., LeCun, Y., and Bregler, C. (2014). *Joint training of a convolutional network and a graphical model for human pose estimation.* Paper presented at the Advances in neural information processing systems.

Unal, M., Onat, M., Demetgul, M., and Kucuk, H. (2014). Fault diagnosis of rolling bearings using a genetic algorithm optimized neural network. *Measurement, 58*, 187-196.

Vachtsevanos, G., and Wang, P. (2001). *Fault prognosis using dynamic wavelet neural networks.* Paper presented at the AUTOTESTCON Proceedings, 2001. IEEE Systems Readiness Technology Conference.

Wang, C.-C., and James Too, G.-P. (2002). Rotating machine fault detection based on HOS and artificial neural networks. *Journal of Intelligent Manufacturing, 13*(4), 283-293. doi:10.1023/a:1016024428793

Wang, C., Zhang, Y., and Zhong, Z. (2008). Fault diagnosis for diesel valve trains based on time–frequency images. *Mechanical Systems and Signal Processing, 22*(8), 1981-1993.

Wang, K. (2014). Intelligent and integrated RFID (II-RFID) system for improving traceability in manufacturing. *Advances in Manufacturing, 2*(2), 106-120.

Wang, K. (2016). Intelligent Predictive Maintenance (IPdM) System–Industry 4.0 Scenario. *WIT Transactions on Engineering Sciences, 113*(1), 259-268.

Wang, K., Li, Z., Braaten, J., and Yu, Q. (2015). Interpretation and compensation of backlash error data in machine centers for intelligent predictive maintenance using ANNs. *Advances in Manufacturing, 3*(2), 97-104.

Wang, K., and Wang, Y. (2012). Towards a next generation of manufacturing: Zero-Defect Manufacturing (ZDM) using data mining approaches: Tapir Academic Press, New York.

Wang, L., Zhang, Z., Long, H., Xu, J., and Liu, R. (2017). Wind turbine gearbox failure identification with deep neural networks. *IEEE Transactions on Industrial Informatics, 13*(3), 1360-1368.

Wang, W. (2007). An adaptive predictor for dynamic system forecasting. *Mechanical*

Systems and Signal Processing, 21(2), 809-823.

Wang, Y., Deng, C., Wu, J., Wang, Y., and Xiong, Y. (2014). A corrective maintenance scheme for engineering equipment. *Engineering Failure Analysis, 36*, 269-283.

Weimer, D., Scholz-Reiter, B., and Shpitalni, M. (2016). Design of deep convolutional neural network architectures for automated feature extraction in industrial inspection. *CIRP Annals-Manufacturing Technology, 65*(1), 417-420.

Wu, T. Y., Chen, J., and Wang, C. (2012). Characterization of gear faults in variable rotating speed using Hilbert-Huang Transform and instantaneous dimensionless frequency normalization. *Mechanical Systems and Signal Processing, 30*, 103-122.

Xiong, G., Shi, D., Chen, J., Zhu, L., and Duan, X. (2013). Divisional fault diagnosis of large-scale power systems based on radial basis function neural network and fuzzy integral. *Electric Power Systems Research, 105*, 9-19.

Xiong, H. Y., Alipanahi, B., Lee, L. J., Bretschneider, H., Merico, D., Yuen, R. K., . . . Hughes, T. R. (2015). The human splicing code reveals new insights into the genetic determinants of disease. *Science, 347*(6218), 1254806.

Yam, R. C. M., Tse, P. W., Li, L., and Tu, P. (2001). Intelligent Predictive Decision Support System for Condition-Based Maintenance. *The International Journal of Advanced Manufacturing Technology, 17*(5), 383-391. doi:10.1007/s001700170173

Yan, J. (2014). *Machinery Prognostics and Prognosis Oriented Maintenance Management*: John Wiley & Sons.

Yang, B.-S., Di, X., and Han, T. (2008). Random forests classifier for machine fault diagnosis. *Journal of mechanical science and technology, 22*(9), 1716-1725.

Yang, B.-S., Oh, M.-S., and Tan, A. C. C. (2008). Machine condition prognosis based on regression trees and one-step-ahead prediction. *Mechanical Systems and Signal Processing, 22*(5), 1179-1193.

Yang, D.-M., Stronach, A., MacConnell, P., and Penman, J. (2002). Third-order spectral techniques for the diagnosis of motor bearing condition using artificial neural networks. *Mechanical Systems and Signal Processing, 16*(2-3), 391-411.

Yang, J., Zhang, Y., and Zhu, Y. (2007). Intelligent fault diagnosis of rolling element bearing based on SVMs and fractal dimension. *Mechanical Systems and Signal Processing, 21*(5), 2012-2024.

Yoon, Y. H., Yoon, E. S., and Chang, K. S. (1992). Process Fault Diagnostics using the Integrated Graph Model. *IFAC Proceedings Volumes, 25*(4), 89-94.

Yu, R. (2000). Theory of Predictive Maintenance and its Application within Framework of Intelligent Control-Maintenance-Management System. *Huazhong University of Science and Technology, Wuhan, China*.

Yuan, M., Wu, Y., and Lin, L. (2016). *Fault diagnosis and remaining useful life estimation of aero engine using lstm neural network.* Paper presented at IEEE International Conference on the Aircraft Utility Systems (AUS).

Zabalza, J., Ren, J., Zheng, J., Zhao, H., Qing, C., Yang, Z., . . . Marshall, S. (2016). Novel segmented stacked autoencoder for effective dimensionality reduction and feature extraction in hyperspectral imaging. *Neurocomputing, 185*, 1-10.

Zamfirescu, C. B., Pirvu, B. C., Schlick, J., and Zuehlke, D. (2013). Preliminary insides for an anthropocentric cyber-physical reference architecture of the smart factory. *Studies in Informatics and Control, 22*(3), 269-278.

Zarei, J., Tajeddini, M. A., and Karimi, H. R. (2014). Vibration analysis for bearing fault detection and classification using an intelligent filter. *Mechatronics, 24*(2), 151-157.

Zhang, L., Yang, F., Zhang, Y. D., and Zhu, Y. J. (2016). *Road crack detection using deep convolutional neural network.* Paper presented at 2016 IEEE International Conference on the Image Processing (ICIP).

Zhang, W., Li, C., Peng, G., Chen, Y., and Zhang, Z. (2018). A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load. *Mechanical systems and signal processing, 100*, 439-453.

Zhang, Z., Wang, Y., and Wang, K. (2013). Fault diagnosis and prognosis using wavelet packet decomposition, Fourier transform and artificial neural network. *Journal of*

Intelligent Manufacturing, 24(6), 1213-1227.

Zhao, J.-l., and Zhao, J.-f. *Application of BPNN and CBR on Fault Diagnosis for Missile Electronic Command System.* (2006). Paper presented at the Sixth World Congress on the Intelligent Control and Automation.

Zhao, R., Wang, J., Yan, R., and Mao, K. (2016). *Machine health monitoring with LSTM networks.* Paper presented at 2016 10th International Conference on the Sensing Technology (ICST).

Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., and Gao, R. X. (2016). Deep Learning and Its Applications to Machine Health Monitoring: A Survey. *arXiv preprint arXiv:1612.07640*.

Zhou, F., Gao, Y., and Wen, C. (2017). A Novel Multimode Fault Classification Method Based on Deep Learning. *Journal of Control Science and Engineering, 2017*.

Zhou, S., Chen, Q., and Wang, X. (2010). *Discriminative Deep Belief networks for image classification.* Paper presented at the 2010 IEEE International Conference on Image Processing.

Ziyan, W., Crossman, J., Cardillo, J., and Murphey, Y. L. (2003, 20-24 July 2003). *Case-base reasoning in vehicle fault diagnostics.* Paper presented at the Proceedings of the International Joint Conference on Neural Networks, 2003.

Zou, A., Deng, R., Mei, Q., and Zou, L. (2017). Fault diagnosis of a transformer based on polynomial neural networks. *Cluster Computing*, 1-9.

Chapter 3
Framework of predictive maintenance in Industry 4.0 concept

3.1 Introduction

The goal of predictive maintenance is to reduce the downtime and the cost of maintenance under the premise of zero failure manufacturing through monitoring the working condition of equipment and predicting the occurrence of failures. Prediction for potential faults allows maintenance to be planned before the fault happens. Ideally, maintenance schedule can be optimized to minimize the cost of maintenance and achieve zero failure manufacturing. However, it is difficult to realize all the advantages of predictive maintenance without the foundation of correlation techniques such as methods to discover fault information, access and integration of industrial big data, and cloud-computing for information sharing. Actually, many manufacturing systems still have not the ability to use multiple data sources to extract relevant information and manage big data due to the high demands on data access and data quality [Jay Lee et al., 2014].

This chapter aims to provide empirical knowledge on the potentials of predictive maintenance in the Industry 4.0 era, and establish a framework for predictive maintenance concerning the Industry 4.0 concept. It may help academics and practitioners to identify and prioritize their steps towards predictive maintenance and condition-based maintenance management under the environment of Industry 4.0. The target of the framework is to minimize the number of unnecessary maintenance performance and leverage the remaining useful life of equipment as much as possible under the premise of zero failure manufacturing.

3.2 Current challenges and potentials of predictive maintenance

3.2.1 Challenges of predictive maintenance

Predictive maintenance measures parameters which can represent the condition of equipment, and carries out the appropriate tasks to optimize the service life of machines and processes without increasing the risk of failure [Garcia et al., 2006]. Comparing with other maintenance strategies, predictive maintenance has following advantages: equipment that requires maintenance is shut down only before imminent failure; reducing the total time spent maintaining equipment; reducing maintenance costs by avoiding catastrophic damage; increasing availability and reliability of machines; extending life of equipment and processes. However, predictive maintenance is facing several challenges as following:

- High demands on data access, data quality and data fusion from multiple sources for data sharing and data publication. Since these sources of data often operate in a heterogeneous environment, integration between the systems is problematic [Aljumaili et al., 2015].
- The capability to deal with industrial big data. To leverage big data, industrial businesses need the ability to support different types of information, the infrastructure to store massive data sets, and the flexibility to leverage the information once collected and stored. In other words, enabling historical analysis of critical trends to enable real-time predictive analysis [J. Liu et al., 2007].
- The prediction accuracy for predictive maintenance. The inaccurate predictive information may result in either unnecessary maintenance, such as early replacement of components, or production downtime because of unexpected machine failures. Therefore, the accuracy of remaining useful life prediction, particularly the long-term prediction, which gives sufficient time to prepare for a maintenance operation, plays an essential role in the full realization of the potentials of predictive maintenance [J. Liu et al., 2007].

3.2.2 From the perspective of maintenance towards Industry 4.0

Industry 4.0 is the superposition of several technological developments related to CPS, IoT, IoS and DM. CPS refers to a new generation of systems with integrated computational and physical capabilities that can interact with humans through many new modalities. The key is the ability to interact with, and expand the capabilities of, the physical world through computation, communication, and control [Baheti and Gill, 2011]. IoT is defined as the ubiquitous access to entities on the internet for the extension of the physical world through a variety of sensing, detection, identification, location tracking and monitoring equipment [Chaves and Nochta, 2011]. IoS pursuits a similar approach with services instead of physical entities. The integration of these developments promotes the cooperation among the partners in the entire system. From the perspective of predictive maintenance, Industry 4.0 accelerates and encourages several developments, which will be reviewed and discussed in this section.

3.2.2.1 Cloud computing environment

Industry 4.0 includes the increasing impact of information and communication technologies on industrial production processes. As one of the driving forces behind Industry 4.0, cloud computing has rapidly emerged as an accepted computing paradigm in many enterprises worldwide due to its flexibility and many other advantages [Bughin et al., 2010]. It can manage shared data from multiple sources efficiently and flexibly in a self-service way and provide a unified service delivery platform for the IoT applications [L. Li et al., 2012]. A cloud-based system provides the technological basis for the provision of data and allows not only creation of community-type services but also building of an open service platform environment which may have features of interactive, collaborative and customizable on demand. From the viewpoint of predictive maintenance, cloud computing environment can efficiently support various smart services and solve several issues such as the memory capacity of equipment, computing power of processor, data security and data fusion from multiple sources.

3.2.2.2 Industrial big data environment

Under the Industry 4.0 era, another significant development is the combination of intelligent analytics and control systems for achieving a new type of manufacturing management and factory transformation. The trend behand this combination is the environment of industrial big data. While big data offers a great potential for revolutionizing all aspects of our society, harvesting of valuable knowledge from big data is not an ordinary task. The large and rapidly growing body of information hidden in the unprecedented volumes of non-traditional data requires both the development of advanced technologies and interdisciplinary teams working in close collaboration [X.-W. Chen and Lin, 2014]. As the most important trend in the development of ICT, deep learning and DM from big data are employed widely to leverage the predictive power in fields like search engines [X.-W. Chen and Lin, 2014], biology [H. Y. Xiong et al., 2015], and astronomy [Jordan and Mitchell, 2015].

The application of industrial big data has been demonstrated in different areas of manufacturing, including production, supply chain, maintenance, quality management, and energy [O'Donovan et al., 2015]. In the research fields of fault diagnosis and prognosis, appropriate sensor installations and various signals can be applied to monitor the working condition of the equipment. In addition, through comparing the current and historical data with data mining approaches, the degree of potential faults, the reliability of certain components and other useful information could be harvested.

3.2.2.3 Smart Factory

Industry 4.0 also facilitates the vision and implementation of "Smart Factory", in which CPS monitor physical processes, create the virtual copies to represent the physical world and make decentralized decisions. Products may not only provide their identity but also record their properties, history and status via ICT technologies. Over cloud computing and distributed control systems, CPS communicate and cooperate with all the available resources. Via the internet of services, both internal and cross-organizational services are offered and utilized to make machines self-aware and actively prevent potential performance issues [Hermann et al., 2015].

From the perspective of maintenance, a smart and self-aware machine system can self-assess its own health and degradation, and further use the fault information from other peers for smart maintenance decisions to avoid potential faults [Jay Lee et al., 2014]. To achieve such intelligence, smart analytics might be used at the individual machine or fleet levels. For a mechanical system, self-aware means the capability to assess the current, past or future working condition of a machine, and output the evaluation result. Such health assessment can be performed through data mining technologies to discover the information collected from the given machine and its ambient environment.

3.3 Structure of the framework

Based on these considerations, a framework for predictive maintenance concerning the Industry 4.0 concept is established. Figure 3.1 shows the general structure of the framework. The target is to minimize the number of unnecessary maintenance performance, leverage the remaining useful life of equipment, and reach zero failure manufacturing through fault diagnosis and accurate prediction for failures and degradation.

The benefits of the framework can be concluded as following;

- It provides an overall understanding and helpful guidance for researchers and practitioners to implement predictive maintenance in the fourth industry revolution.
- It can collect and leverage all available data and information from types of data resources such as mounted sensors, control systems, or cloud database to evaluate the working condition.
- Intelligent fault diagnosis and prognosis can be made to detect when, where, which equipment and which component may have impending failures.
- Faults and degradation could be accurately predicted and assessed through deep learning models for maintenance scheduling optimization.
- Zero failure performance and subsequent zero defect manufacturing can be reached since all the potential faults could be predicted and fixed before they occur.
- It can make predictive maintenance decision to prevent occurrence and development of failures effectively, ensure the safety of equipment, and reduce the total cost of maintenance by minimizing the number of unnecessary maintenance performance.

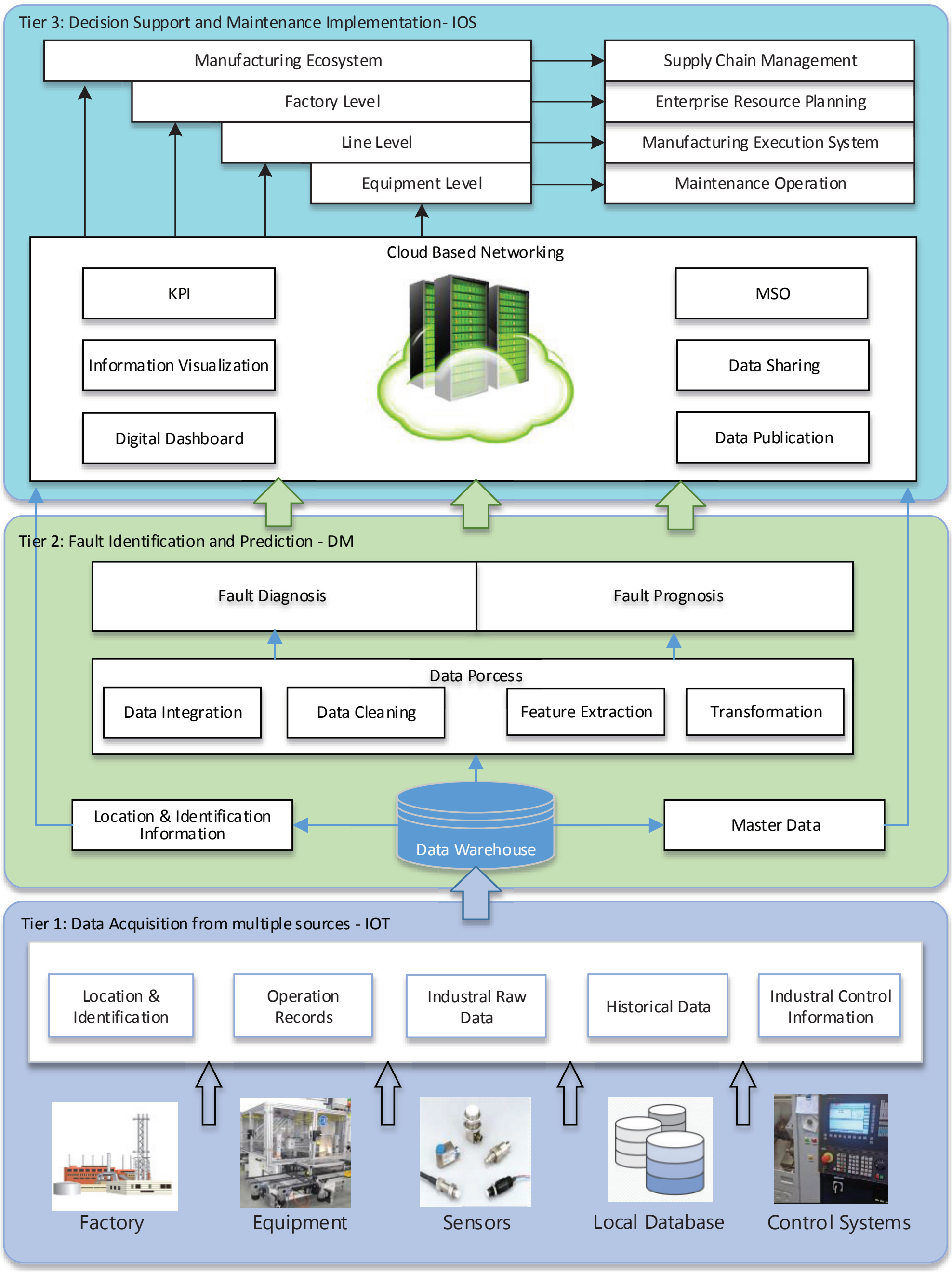

Figure 3.1 Framework of predictive maintenance in Industry 4.0 concept

Figure 3.1 illustrates the general structure of the proposed framework, which consists of three tiers. The first tier is data acquisition from multiple sources, which allows all relevant devices and data sources interconnect with each other. Fault identification and prediction is the second tier. It is responsible to discover the fault information from collected data. The third tier is decision support and maintenance implementation, which is in charge of presenting information or knowledge obtained from data mining, providing optimized maintenance scheduling and finally interacting with the physical world according to the computation results in cyber world. Systems based on this framework can monitor plant floor assets, link the production and maintenance operations systems, obtain data, collect feedback from a remote customer site, and integrate it into upper-level enterprise applications, discovery hidden information about impending failures, and generate maintenance knowledge. It can also monitor the state of manufacturing processes and predict the condition of the equipment. Systems based on this framework can make a maintenance decision to prevent the occurrence and development of failures effectively, ensure equipment and personal safety, and reduce the economic loss caused by failures. It can use fault diagnosis, performance assessment of the degrading level, and fault prognosis models to achieve near-zero-breakdown performance and improve the productivity of a company. The techniques of the three tiers are elaborated in the following sections.

3.4 Tier 1: Data acquisition from multiple sources

As the first tier in the framework, data acquisition from multiple sources is the physical foundation of predictive maintenance. It is based on IoT technologies and plays the same role as IoT in Industry 4.0, the physical access to “things or objects”. The main task of this tier is selecting suitable sensors, data sources and data collection strategy to extend the physical world through a variety of sensing, detection, identification and connect the objects or make them interact with each other. The data acquisition process transforms the collected data into domains that are of the most information to represent the working condition of equipment or fusion of several domains. The most important factor in this tier is the ability of virtualization, which means to monitor the physical processes and create a virtual copy of the physical world by linking the collected data to virtual or simulation models. The selection of the suitable data sources is the key of the effectiveness in condition-based predictive maintenance, and a complete data acquisition system can improve the ability of virtualization, subsequent correction and efficiency of diagnosis and prognosis directly. Four common types of data sources, which can be leveraged to achieve predictive maintenance in Industry 4.0 concept are listed as following.

3.4.1 Sensors

The rapid advancement in sensor technology has made streaming real-time data easier than ever. Various sensors, such as micro-sensors, ultrasonic, vibration, and acoustic emission sensors, can be designed to collect various types of data and generate rich sources of industrial data. The selection of sensors decides the representation of the machine health by the collected data, considering both the specifications and cost-effectiveness. Actually, how to achieve smart sensors is also a significant research field in condition monitoring [Son et al., 2009]. For condition based maintenance or monitoring, we usually select the sensors based on the performance in representing the physical characteristics. A variety of sensor systems or transducers exists and can be applied for effectively monitoring various process parameters [K. Wang, 2003]. Some popular sensor systems or transducers widely applied in condition monitoring are listed as follow:

- Mechanical sensor systems to monitor parameters such as acceleration, displacement, velocity, torque, location, strain, and cutting forces (static and dynamic).
- Optical transduces such as photo detectors, and lasers.
- Thermal transducers such as thermocouples, and thermography.
- Audible sensors such as ultrasonic sensors, and acoustic emission sensors
- Environment sensors systems such as the spectrometer, PH indicators, and temperature sensors.

In predictive maintenance, one important factor to select the suitable type of sensors is the expected preventing time. For example, vibration sensors are widely leveraged for predicting and detecting early failures in mechanical systems and manufacturing processes, since the preventing time of vibration sensors may be months. The preventing time for audible sensors may be several weeks and thermal sensors may be a few days. Another important factor is based on the physical characteristics. As discussed above, the essence of condition-based maintenance is to detect the signs of impending failures, and predict the potential faults or degradation based on these signs. Therefore, to select the signals that can represent the physical characteristics of the signs of impending failures is also an important principal for selection. In most cases, it may depend on the empirical knowledge in relevant domains.

3.4.2 Industrial control systems

In Industry 4.0 concept, another significant data source is from inherent industrial control systems like supervisory control and data acquisition (SCADA). Those systems are normally combined with data acquisition systems through adding the use of coded signals over communication channels. Information about the status of both local or remote equipment is usually recorded in these systems [Desai et al., 2014]. Therefore, many precious and necessary information can be obtained directly from these industrial control systems to the data warehouse. Actually, a lot of manufacturing factories or enterprises have been gathering data in kinds of industrial control systems for many years, which means tremendous historical information may be harvested. This advancement benefits from the implementation of IoT directly.

3.4.3 Cloud database

Another benefit form IoT is the wide application of cloud database. Today's manufacturing involves all activities ranging from product design, production, fabrication, testing, maintenance and all other stages of a product life cycle [B.-H. Li et al., 2011]. An enabling factor in becoming an agile manufacturer or smart factory has been the development of manufacturing support technology that allows the marketers, designers and production personnel to leverage a common cloud database. The database could be used to share data or information about production capacities and problems, particularly where small initial problems may have larger downstream effects [Zhang et al., 2014]. Access to the cloud database means acquirement of all relevant information about those activities.

3.4.4 Location & identification

Information about location and identification may come from various types of devices or systems, such as geographic information system, mobile devices, RFID or GPS electronic devices. Among these technologies, RFID is a rapidly developing technology, which relies on wireless communication for automatic identification of things or objects. It has been widely adopted in supply chain management and manufacturing industry for the purpose of automatic identification and tracking of objects [Nikitin and Kelly, 2014]. A typical RFID system consists of tags, reader, and antenna. A RFID reader could communicate with the tags in electromagnetic field. Whenever a tag enters the interrogation region, it can be detected, located, and identified by the RFID reader [Alarifi et al., 2016]. For predictive maintenance, both the equipment and its main components can be labelled by RFID tags for location and identification.

3.5 Tier 2: Fault identification and prediction

The second tier is fault identification and prediction, which can be considered as the core processor for data mining in the framework. The function is to discover knowledge or information relevant to faults or degradation from data. All the data collected in the first tier will be stored in the data warehouse for further data mining in the second tier. However, during the process of knowledge discovery, if there is too much irrelevant and redundant information, like noise or unreliable data, then it would be more challenging during the training phase. Therefore, it is necessary to process data before extracting information from data (This step may also be called as data preprocess in some literatures). After data process, diagnosis and prognosis models can be established for fault identification and prediction. Figure 3.2 shows the process flow of data analysis for fault diagnosis and prognosis in this tier. The diagnosis and prognosis methods encouraged in this thesis are based on deep learning approaches, which could ideally solve some challenges during the implementation of predictive maintenance. In Chapter 2, the superiorities of several deep learning architectures have been discussed, which could be used as guidelines to select suitable approaches in this tier.

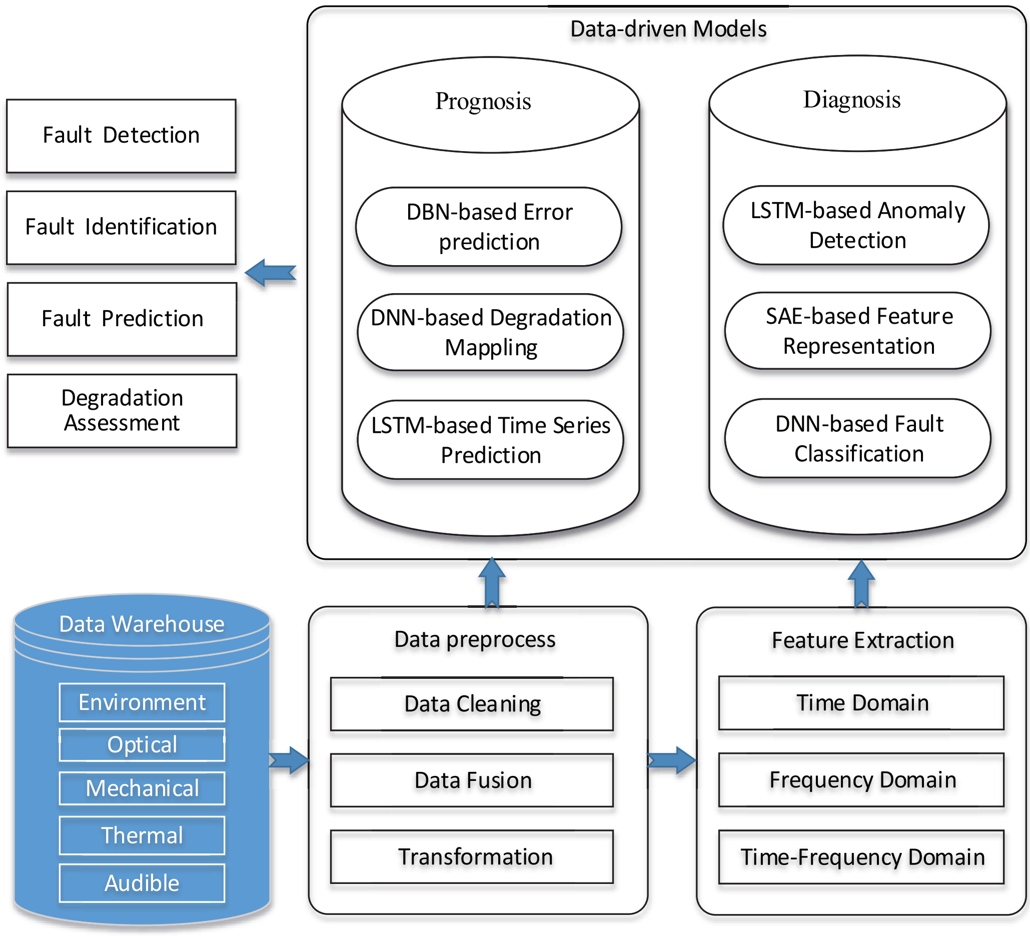

Figure 3.2 Data analysis process for fault diagnosis and prognosis

3.5.1 Data preprocess

The major steps involved in data preprocessing includes data cleaning, data fusion, data reduction, data transformation, and feature extraction. Data cleaning is the process of detecting and correcting corrupt or inaccurate records from the database by filling in missing values, smoothing noisy data, identifying or removing outliers, and resolving inconsistencies. Data fusion is the process of merging data from multiple data stores. Careful integration can help to reduce and avoid redundancies and inconsistencies in the resulting data set. In data transformation, the data are transformed or consolidated into forms appropriate for knowledge discovery, so that the data mining process may be more efficient, and the patterns found may be easier to understand.

Data reduction obtains a reduced representation of the data set that is much smaller in volume, and can produce the same (or almost the same) analytical results. There are many dimensionality reduction methods. Among them, a straightforward approach is to apply feature extraction methods to the data set, which extracts features from preprocessed signals that are characteristic of an incipient failure or fault. Generally, the features can be extracted from three domains: time domain, frequency domain and time-frequency domain. In Chapter 2, a brief review of all popular and widely applied feature extraction methods for fault diagnosis and prognosis based on literatures has been listed in Table 2.1. In Chapter 6, a novel application of deep learning-based representation learning will also be introduced and leveraged in an experiment to reduce the size of multiple features sequence through SAE.

In addition, the development of the storage media and computation ability produces massive data during the data acquisition process. Although deep learning methods have the ability to deal with industrial raw data, it may increase the required complexity and computation load of the data-driven models. Data preprocessing can effectively clean raw data, reduce dimension of data, and store it back to the warehouse for knowledge discovery. Massive data can be converted to features or statistical values as the input variables for diagnosis and prognosis models. Therefore, suitable data preprocess operations could enhance the signal characteristics and eventually facilitate the efficient extraction of useful information in both faults diagnosis and prognosis.

3.5.2 Data mining for fault diagnosis and prognosis

Data mining has the capability to discover hidden links, recognize unknown patterns, and predict future trends by digging through and analysing enormous sets of data [Sumathi and Sivanandam, 2006]. The functions, or models, of data mining can be categorized according to the task performed [Siguenza-Guzman et al., 2015], such as clustering, classification, decision trees, predication, regression, association, etc. Normally, the analysis methods of data mining can be categorized into two groups: statistics, and AI [Girija, 2006].

A statistical model is a set of mathematical functions, which describe the behaviour of the objects in a target class in terms of random variables and their associated probability distributions [Han et al., 2011]. Due to the fact that statistics has an inherent connection with data mining during data collection, analysis, interpretation, and presentation, it is widely leveraged to model data and data classes during the process of data mining. Examples of statistics includes regression analysis, cluster analysis, and discriminate analysis, etc. However, since most statistical models are based on probability distributions, they usually focus on the reliability or probability analysis of equipment failure instead of precisely fault prediction or degradation assessment, which is important to predictive maintenance.

The second family root of data mining is AI, which is built upon heuristic algorithm. It includes several techniques such as genetic algorithm, artificial neural network, FLS. The main idea is to apply human-thought-like processing to solve problems. It uses techniques for writing computer code to represent and manipulate knowledge, which exactly fits in the computer processing in modern business environment [Girija, 2006]. As the most important approach in AI [Michalski et al., 2013], machine learning investigates how computers can study and make predictions based on data [Kohavi and Provost, 1998]. Machine learning is employed in a range of computing tasks to learn to recognize complex patterns and make appropriate decisions automatically. It usually can be divided into two main types: predictive or supervised learning, and descriptive or unsupervised learning [Murphy, 2012]. In the predictive or supervised learning process, the goal is to form a mapping from inputs x to outputs y, given a labeled set of input-output data $D = \{(x_i|y_i)\}\ i = 1,2,3 \dots N$. Here D is the training set, and N is the number of training examples. In descriptive or unsupervised learning, the learning process is unsupervised since there are no class labelled in the input samples. Here only the inputs will be given, $D = \{x_i\}\ i = 1,2,3 \dots N$, and the goal is to find interesting patterns and knowledge from large amounts of data. This is a much less well-defined problem, since we are not told what kinds of patterns to look for, and there is no obvious error metric for evaluating the results. Currently, due to the rapid development of deep learning approaches, it provides alternative methods with outstanding performance to deal with the issues about fault prediction and degradation assessment, which were once considered as changeling problems in predictive maintenance. Data mining benefits from these technologies, but differs from the objective pursued: extracting patterns, describing trends, or predicting behaviours. It has been applied in a wide range of domains, where large amounts of data are available for the identification of unknown or hidden information [Siguenza-Guzman et al., 2015]. Data mining approaches in the framework mainly focus on fault detection, identification, prediction, and degradation assessment for predictive maintenance, which could also be divided into fault diagnosis and prognosis.

The aim of fault diagnosis in the framework is to detect abnormal condition of equipment before the failure happens, and identify which type of failures or which component may have impending failures. Fault diagnosis have been developed and found extensive utility in a wide range of application domains in recent years. Typically, it can be divided into two major categories: model based and data-driven [G. J. Vachtsevanos et al., 2006]. Model-based technique depends on the accuracy of dynamic system model. It takes advantage of the actual system and model to generate the difference between the two outputs, which is indicative of an impending fault condition. However, in many manufacturing systems, it is difficult to establish a high-accuracy dynamic system model directly. On the other hand, data-driven techniques often only address anticipated fault condition, where a fault model is a collection of constructs like neural networks and SVM, which must be trained first with known prototype fault patterns. Normally, if the historical data can be obtained easily, the data-driven is very useful to identify the fault and evaluate the working condition. When only part of historical data can be obtained, the hybrid techniques, which combine the data-driven techniques and model-based techniques, can be used to evaluate current conditions of manufacturing systems or the products. The semi-supervised learning method also can be used to evaluate condition and identify fault when only part of historical data is available. It is also very effective and widely used for fault diagnosis. Since fault diagnosis is a subjective problem in nature, the most suitable method usually depends on the practical issue. In Chapter 2, Table 2.2, an overview of all popular AI techniques, which have been widely applied in fault diagnosis in literature are listed. According to literature review, all these methods such as BPNN [Jafar et al., 2010; Rohani et al., 2011], RBFNN [G. Xiong et al., 2013], CPNN [Phillips et al., 2015], CCNN [Phillips et al., 2015], SOM [Rai and Upadhyay, 2017; C.-C. Wang and James Too, 2002], PNN [Zou et al., 2017], dynamic neural network [Abed et al., 2014], SVM [Konar and Chattopadhyay, 2011; Rai and Upadhyay, 2017], DBN [Gan and Wang, 2016; Tamilselvan and Wang, 2013], and deep neural network (DNN) [Jia et al., 2016; L. Wang et al., 2017], both deep learning or conventional machine learning approaches, have proved their superiority or priority in certain fields.

As to prognosis, the conventional AI techniques applied are mainly based on shallow learning algorithms, including SVM [Kim et al., 2012; Rai and Upadhyay, 2017], SOM [Rai and Upadhyay, 2017], DT [Yang et al., 2008], DWNN [G. Vachtsevanos and Wang, 2001], and BPNN [Rohani et al., 2011]. ANN-based machinery prognostics approach was once the most commonly found data-driven technique for fault prognosis [Heng et al., 2009]. Shallow and deep learners are distinguished by the depth of their credit assignment paths, which are chains of possibly learnable, causal links between actions and effects [Schmidhuber, 2015]. The network learns the unknown function by adjusting its weights with repetitive observations of inputs and outputs. Numerous studies across various disciplines have demonstrated the merits of ANNs, including the abilities to perform faster than system identification techniques in multivariate prognosis [J Lee, 2007] and perform at least as good as the best traditional statistical methods, without requiring untenable distributional assumptions [Joshi and Reeves, 2006].

As reported in [Heng et al., 2009], many data-driven methods such as artificial neural network are capable in modelling complex phenomenon, but they may require more complex structures to represent the phenomenon with high complexity. In predictive maintenance, forecasting potential failures in future is more challenging than detection or identification of occurred or impending failures since the data which can represent the working condition in the future is absent. In order to compensate the absence of target condition, more comprehensive and detailed history data and data-driven model with higher complexity and generalization are required to track or evaluate the development of certain faults and degradations. Therefore, issues in prognosis are usually with higher complexity than in diagnosis. However, if we intentionally increase the complexity of a typical artificial neural network through raising the number of hidden layers, it will suffer from the now famous problem of vanishing or exploding gradients, which is a widely known limitation of conventional ANNs. This is the reason why accurate prediction of potential faults in future has been a hot issue and challenge in predictive maintenance for decades. In recent years, as a latest and advanced research field, deep artificial neural networks have accelerated its application and shown their superiorities in predictive maintenance [Gan and Wang, 2016]. Deep learning is making major advances in solving problems that have resisted the best attempts of the AI community for many years. By increasing the complexity of data-driven models themselves and some special architectures such as energy-based activation functions, deep learning enables data-driven models the ability of self-learning through training data. It has turned out to be very good at discovering intricate structures in high-dimensional data and is therefore applicable to many domains of science, business and government [LeCun et al., 2015]. Currently, various deep learning algorithms, such as deep belief networks [Gan and Wang, 2016; Tamilselvan and Wang, 2013] and deep neural networks [L. Wang et al., 2017], have been applied successfully in predictive maintenance. A systematic research about several deep learning approaches applied for predictive maintenance along with their superiorities has been discussed in Chapter 2. In Chapter 5 and Chapter 6, some novel applications such as DBN-based error prediction, DNN-based degradation assessment, SAE-based multiple feature sequence reconstruction, and LSTM-based anomaly detection will also be introduced and detailed through practical applications.

3.6 Tier 3: Decision support and maintenance implementation

The third tier is decision support and maintenance implementation. The function is to manipulate the analysis results in data mining tier, transform them into meaningful information or knowledge for maintenance strategy, share and publish these information on a common cloud based networking, and eventually provide optimal schedule for maintenance implementation. In one word, this tier is the infrastructure for services related to maintenance via Internet. The main functions in this tier are information visualization, maintenance scheduling optimization, and the interoperation.

3.6.1 Information visualization

As the study of transforming data, information, and knowledge into interactive visual representations, information visualization is significant to users because it provides mental models of information [S. Liu et al., 2014]. The aim is to aid users in exploring, understanding, and analyzing data through progressive, iterative visual exploration [Shiravi et al., 2012]. In recent years, with the development of visualization technology and boom in big data analytics, more and more experts and scholars have begun to combine visualization technology with condition monitoring, fault diagnosis and prognosis technologies to better solve the shortcomings of traditional methods [Chu et al., 2017]. In predictive maintenance, applying key performance indicators (KPI) is one of the most efficient and popular method to illustrate the result of data mining, since there always exist certain indicators, which can be regarded as the critical ones for equipment or production process in the practical industrial applications [Yin et al., 2016]. Subsequently, KPI tracking can be achieved through publishing on a digital dashboard, and give the entire organization insights into the current condition, performance or certain degradation.

3.6.2 Maintenance scheduling optimization

Maintenance scheduling optimization means deciding which maintenance activities to perform, and when, such that one or several objectives are optimized [Gustavsson et al., 2014]. The target is to make maintenance decision based on current available information to optimize certain objectives such as maintenance cost, and development of potential failures. In predictive maintenance, the major objectives of maintenance scheduling are maximizing equipment up-time under zero-failure manufacturing, minimizing time to repair, and decreasing total maintenance cost according to the prediction or assessment from data mining. Other objectives and constraints, such as minimizing logistics footprint and cost of transportation, may also be included in the optimization scheme if dictated by specific system requirements or logistics network [G. J. Vachtsevanos et al., 2006].

However, in practical industrial applications, attention shall also be paid to the relationship between production and maintenance, which has been considered as a conflict in management decision [Berrichi et al., 2010]. Hence, the issue about how to capture the trade-off among the objective of both production and maintenance shall also be considered during scheduling, which make it as a type of NP problem. In recent years, to solve complex scheduling problems for predictive maintenance, heuristic algorithms such as genetic algorithm [C.-H. Wang and Tsai, 2014], particle swarm optimization [Liao et al., 2011], genetic simulated annealing algorithm [X. Li et al., 2015], imperialist competitive algorithm [Zandieh et al., 2017], artificial bee colony algorithm [Dalfard and Mohammadi, 2012], and ant colony algorithm [Saleh et al., 2017]. Although these techniques may not guarantee global optimal solutions, they are usually not restricted to the size or structure of the problem, and could provide good solutions for maintenance scheduling optimization within reasonable computational time [Doostparast et al., 2014]. Therefore, all these kinds of methods could be selected and employed in the framework for maintenance scheduling optimization. In Chapter 5, a novel HDPS-BPSO maintenance scheduling approach will be proposed and utilized to show the advantage of predictive maintenance compared with preventive one.

3.6.3 Interoperation

In Industry 4.0 companies, the cyber, the physical and humans are connected over IoT and the IoS. Interoperability enables valuable connections, whether across processes, between people and information, or among companies. It offers the ability for systems to understand each other and leverage functionality of others. The word 'inter-operate' implies that one system performs an operation for another system. From the computer technology point of view, it is the faculty for several heterogeneous computer systems to function jointly and to give access to their resources in a reciprocal way [D. Chen et al., 2008]. For predictive maintenance, interoperability refers to the ability to interact in data, services and processes between enterprise systems. After we acquired the detailed information or knowledge about the failures based on the result of data analysis in the cyber world, those information or knowledge shall be employed to interact the physical world, to be more specific, to implement maintenance and evaluate the influence of degradation or failures and solutions in equipment level, line level, factory level, and even manufacturing ecosystem. Therefore, it is important for practitioners to have a consideration about coordination among those systems such as maintenance operation, manufacturing execution system, enterprise resource planning, and supply chain management.

3.7 Summary

The framework of predictive maintenance in Industry 4.0 concept is established to monitor the manufacturing system and process, identify and predict impending or potential failures, and minimize the number of unnecessary maintenance performance under the premise of zero failure manufacturing. Based on this framework, approximate maintenance scheduling could be made through the prediction of impending failures or certain degradation to ensure zero failure manufacturing and minimize the cost of maintenance.

The framework is constructed based on the architecture of CPS. In the first tier, data acquisition from multiple sources, types of data that can represent the conditions in the physical world is transformed into the cyber world through the implementation of IoT. The second tier, diagnosis and prognosis, is responsible to discover the information or knowledge about the failures and degradation through data mining approaches in the cyber world. The third tier, decision support and maintenance implementation, provides types of service based on the information or knowledge obtained from data mining. It can also offers decision support according to the data mining results and eventually interact with the physical world.

References

Abed, W., Sharma, S., and Sutton, R. (2014). *Diagnosis of bearing fault of brushless DC motor based on dynamic neural network and orthogonal fuzzy neighborhood discriminant analysis.* Paper presented at 2014 UKACC International Conference on the Control (CONTROL).

Alarifi, A., Al-Salman, A., Alsaleh, M., Alnafessah, A., Al-Hadhrami, S., Al-Ammar, M. A., and Al-Khalifa, H. S. (2016). Ultra wideband indoor positioning technologies: Analysis and recent advances. *Sensors, 16*(5), 707.

Aljumaili, M., Wandt, K., Karim, R., and Tretten, P. (2015). eMaintenance ontologies for data quality support. *Journal of Quality in Maintenance Engineering, 21*(3), 358-374.

Baheti, R., and Gill, H. (2011). Cyber-physical systems. *The impact of control technology, 12*, 161-166.

Berrichi, A., Yalaoui, F., Amodeo, L., and Mezghiche, M. (2010). Bi-objective ant colony optimization approach to optimize production and maintenance scheduling. *Computers & Operations Research, 37*(9), 1584-1596.

Bughin, J., Chui, M., and Manyika, J. (2010). Clouds, big data, and smart assets: Ten tech-enabled business trends to watch. *McKinsey Quarterly, 56*(1), 75-86.

Chaves, L. W., and Nochta, Z. (2011). Breakthrough towards the internet of things *Unique Radio Innovation for the 21st Century* (pp. 25-38): Springer.

Chen, D., Doumeingts, G., and Vernadat, F. (2008). Architectures for enterprise integration and interoperability: Past, present and future. *Computers in Industry, 59*(7), 647-659.

Chen, X.-W., and Lin, X. (2014). Big data deep learning: challenges and perspectives. *IEEE access, 2*, 514-525.

Chu, S., Zhang, Q., Xu, B., Liu, F., and Liang, L. (2017). *Visualization method for spectrum analysis of vibration acceleration signals.* Paper presented at 2017 IEEE International Conference on the Prognostics and Health Management (ICPHM).

Dalfard, V. M., and Mohammadi, G. (2012). Two meta-heuristic algorithms for solving multi-objective flexible job-shop scheduling with parallel machine and maintenance constraints. *Computers & Mathematics with Applications, 64*(6), 2111-2117.

Desai, P., Mahale, S., Desai, P., and Karamchnadani, S. (2014). Smart SCADA and automation system in power plants. *Int. J. Curr. Eng. Technol*.

Doostparast, M., Kolahan, F., and Doostparast, M. (2014). A reliability-based approach to optimize preventive maintenance scheduling for coherent systems. *Reliability Engineering & System Safety, 126*, 98-106.

Gan, M., and Wang, C. (2016). Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings. *Mechanical Systems and Signal Processing, 72*, 92-104.

Garcia, M. C., Sanz-Bobi, M. A., and del Pico, J. (2006). SIMAP: Intelligent System for Predictive Maintenance: Application to the health condition monitoring of a windturbine gearbox. *Computers in Industry, 57*(6), 552-568.

Girija, N., & Srivatsa, S.K. (2006). A research study: Using data mining in knowledge base business strategies. *Information Technology Journal, 5(3)*, 590–600.

Gustavsson, E., Patriksson, M., Strömberg, A.-B., Wojciechowski, A., and Önnheim, M. (2014). Preventive maintenance scheduling of multi-component systems with interval costs. *Computers & Industrial Engineering, 76*, 390-400.

Han, J., Kamber, M., and Pei, J. (2011). *Data mining: concepts and techniques*: Elsevier.

Heng, A., Zhang, S., Tan, A. C., and Mathew, J. (2009). Rotating machinery prognostics: State of the art, challenges and opportunities. *Mechanical Systems and Signal Processing, 23*(3), 724-739.

Hermann, M., Pentek, T., and Otto, B. (2015). Design principles for Industrie 4.0 scenarios: a literature review. *Technische Universität Dortmund, Dortmund*.

Jafar, R., Shahrour, I., and Juran, I. (2010). Application of Artificial Neural Networks (ANN) to model the failure of urban water mains. *Mathematical and Computer Modelling, 51*(9), 1170-1180.

Jia, F., Lei, Y., Lin, J., Zhou, X., and Lu, N. (2016). Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. *Mechanical Systems and Signal Processing, 72*, 303-315.

Jordan, M. I., and Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science, 349*(6245), 255-260.

Joshi, R., and Reeves, C. (2006). *Beyond the Cox model: artificial neural networks for survival analysis part II.* Paper presented at the Proceedings of the eighteenth international conference on systems engineering.

Kim, H.-E., Tan, A. C., Mathew, J., and Choi, B.-K. (2012). Bearing fault prognosis based on health state probability estimation. *Expert systems with applications, 39*(5), 5200-5213.

Kohavi, R., and Provost, F. (1998). Glossary of terms. *Machine learning, 30*(2-3), 271-274.

Konar, P., and Chattopadhyay, P. (2011). Bearing fault detection of induction motor using wavelet and Support Vector Machines (SVMs). *Applied Soft Computing, 11*(6), 4203-4211.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436-444.

Lee, J. (2007). *A systematic approach for developing and deploying advanced prognostics technologies and tools: methodology and applications.* Paper presented at the Proceedings of the Second World Congress on Engineering Asset Management, Harrogate, UK.

Lee, J., Kao, H.-A., and Yang, S. (2014). Service innovation and smart analytics for industry 4.0 and big data environment. *Procedia CIRP, 16*, 3-8.

Li, B.-H., Zhang, L., Ren, L., Chai, X.-D., Tao, F., Luo, Y.-L., . . . Zhao, X. (2011). Further discussion on cloud manufacturing. *Computer Integrated Manufacturing Systems, 17*(3), 449-457.

Li, L., Xinrui, L., and Xinyu, L. (2012). Cloud-Based Service Composition Architecture for Internet of Things. *Communications in Computer and Information Science, Springer, 312*, 559-564.

Li, X., Wen, J., Zhou, R., and Hu, Y. (2015). *Study on Resource Scheduling Method of Predictive Maintenance for Equipment Based on Knowledge.* Paper presented at 2015 10th International Conference on the Intelligent Systems and Knowledge Engineering (ISKE).

Liao, R., Zheng, H., Grzybowski, S., and Yang, L. (2011). Particle swarm optimization-least squares support vector regression based forecasting model on dissolved gases in oil-filled power transformers. *Electric Power Systems Research, 81*(12), 2074-2080.

Liu, J., Djurdjanovic, D., Ni, J., Casoetto, N., and Lee, J. (2007). Similarity based method for manufacturing process performance prediction and diagnosis. *Computers in Industry, 58*(6), 558-566. doi:http://dx.doi.org/10.1016/j.compind.2006.12.004

Liu, S., Cui, W., Wu, Y., and Liu, M. (2014). A survey on information visualization: recent advances and challenges. *The Visual Computer, 30*(12), 1373-1393.

Michalski, R. S., Carbonell, J. G., and Mitchell, T. M. (2013). *Machine learning: An artificial intelligence approach*: Springer Science & Business Media.

Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*: MIT press.

Nikitin, P., and Kelly, S. J. (2014). Modulation index (depth) based grouping, addressing, and fingerprinting rfid tags: Google Patents.

O'Donovan, P., Leahy, K., Bruton, K., and O'Sullivan, D. T. (2015). Big data in manufacturing: a systematic mapping study. *Journal of Big Data, 2*(1), 1-22.

Phillips, J., Cripps, E., Lau, J. W., and Hodkiewicz, M. (2015). Classifying machinery condition using oil samples and binary logistic regression. *Mechanical Systems and Signal Processing, 60*, 316-325.

Rai, A., and Upadhyay, S. H. (2017). Intelligent bearing performance degradation assessment and remaining useful life prediction based on self-organising map and support vector regression. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 0954406217700180.

Rohani, A., Abbaspour-Fard, M. H., and Abdolahpour, S. (2011). Prediction of tractor repair and maintenance costs using artificial neural network. *Expert systems with applications, 38*(7), 8999-9007.

Saleh, N., Rosati, S., Sharawi, A., Wahed, M. A., and Balestra, G. (2017). An Optimal Scheduling for Medical Equipment Preventive Maintenance Over a Finite Planning Horizon Using Ant Colony Algorithm. *Journal of Clinical Engineering, 42*(3), 142-149.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks, 61*, 85-117.

Shiravi, H., Shiravi, A., and Ghorbani, A. A. (2012). A survey of visualization systems for network security. *IEEE Transactions on visualization and computer graphics, 18*(8), 1313-1329.

Siguenza-Guzman, L., Saquicela, V., Avila-Ordóñez, E., Vandewalle, J., and Cattrysse, D. (2015). Literature review of data mining applications in academic libraries. *The Journal of Academic Librarianship, 41*(4), 499-510.

Son, J.-D., Niu, G., Yang, B.-S., Hwang, D.-H., and Kang, D.-S. (2009). Development of smart sensors system for machine fault diagnosis. *Expert systems with applications, 36*(9), 11981-11991.

Sumathi, S., and Sivanandam, S. (2006). *Introduction to data mining and its applications* (Vol. 29): Springer.

Tamilselvan, P., and Wang, P. (2013). Failure diagnosis using deep belief learning based health state classification. *Reliability Engineering & System Safety, 115*, 124-135.

Vachtsevanos, G., and Wang, P. (2001). *Fault prognosis using dynamic wavelet neural networks.* Paper presented at the AUTOTESTCON Proceedings, 2001. IEEE Systems Readiness Technology Conference.

Vachtsevanos, G. J., Lewis, F., Hess, A., and Wu, B. (2006). *Intelligent fault diagnosis and prognosis for engineering systems*: Wiley Online Library.

Wang, C.-C., and James Too, G.-P. (2002). Rotating machine fault detection based on HOS and artificial neural networks. *Journal of Intelligent Manufacturing, 13*(4), 283-293. doi:10.1023/a:1016024428793

Wang, C.-H., and Tsai, S.-W. (2014). Optimizing bi-objective imperfect preventive maintenance model for series-parallel system using established hybrid genetic algorithm. *Journal of Intelligent Manufacturing, 25*(3), 603-616.

Wang, K. (2003). *Intelligent condition monitoring and diagnosis systems: a computational intelligence approach* (Vol. 93): IOS press.

Wang, L., Zhang, Z., Long, H., Xu, J., and Liu, R. (2017). Wind turbine gearbox failure identification with deep neural networks. *IEEE Transactions on Industrial Informatics, 13*(3), 1360-1368.

Xiong, G., Shi, D., Chen, J., Zhu, L., and Duan, X. (2013). Divisional fault diagnosis of large-scale power systems based on radial basis function neural network and fuzzy integral. *Electric Power Systems Research, 105*, 9-19.

Xiong, H. Y., Alipanahi, B., Lee, L. J., Bretschneider, H., Merico, D., Yuen, R. K., . . . Hughes, T. R. (2015). The human splicing code reveals new insights into the genetic determinants of disease. *Science, 347*(6218), 1254806.

Yang, B.-S., Oh, M.-S., and Tan, A. C. C. (2008). Machine condition prognosis based on regression trees and one-step-ahead prediction. *Mechanical Systems and Signal Processing, 22*(5), 1179-1193.

Yin, S., Ding, S. X., and Zhou, D. (2016). Diagnosis and prognosis for complicated industrial systems—Part I. *IEEE Transactions on Industrial Electronics, 63*(4), 2501-2505.

Zandieh, M., Khatami, A., and Rahmati, S. H. A. (2017). Flexible job shop scheduling under condition-based maintenance: Improved version of imperialist competitive algorithm. *Applied Soft Computing, 58*, 449-464.

Zhang, L., Luo, Y., Tao, F., Li, B. H., Ren, L., Zhang, X., . . . Liu, Y. (2014). Cloud manufacturing: a new manufacturing paradigm. *Enterprise Information Systems, 8*(2), 167-187.

Zou, A., Deng, R., Mei, Q., and Zou, L. (2017). Fault diagnosis of a transformer based on polynomial neural networks. *Cluster Computing*, 1-9.

Chapter 4
Implementation of predictive maintenance in machining centers

4.1 Introduction

In this chapter, an Industry 4.0 scenario of predictive maintenance for machining centers will be introduced. This research comes from a project named green monitoring and has been supported by a grant from Norway through the Norwegian Financial Mechanism 2009-2014, in the frame of the Green Industry Innovation Programme Bulgaria. The scenario demonstrated the guideline and implementation of predictive maintenance in machining centers.

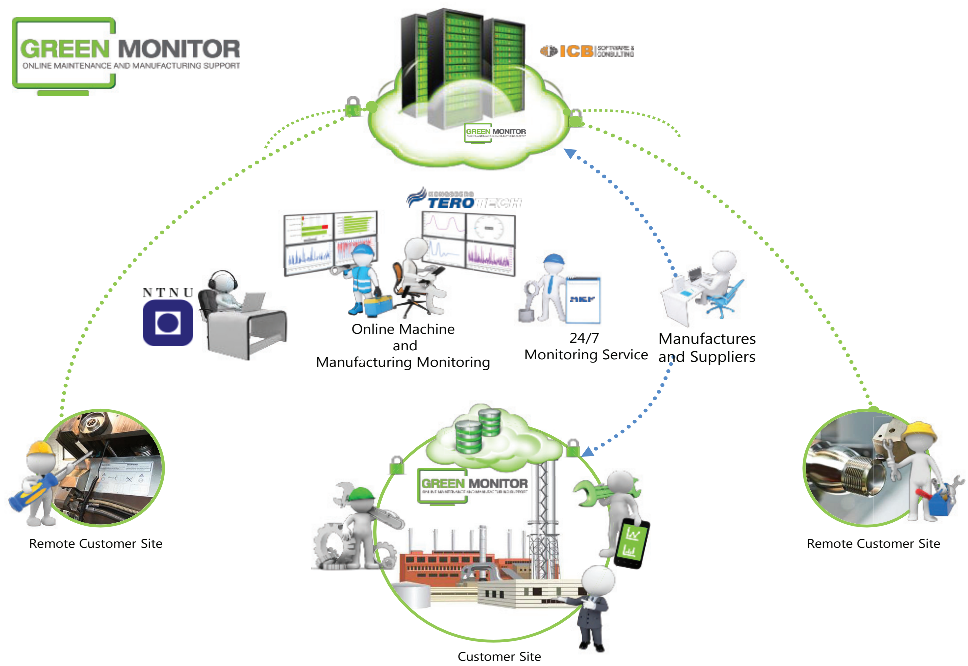

Figure 4.1 Green monitoring system

In this study, Norwegian University of Science and Technology (NTNU) cooperated with InterConsult Bulgaria (ICB) and Kongsberg Terotech (KTT), who provided their assistance in the software, empirical knowledge, and data sources for the study. The aim is to decrease the cost on faults, defects and maintenance during the manufacturing process through remote condition monitoring and data mining. As shown in Figure 4.1, the data is acquired from remote customer site and transported to the data mining center for data analysis. After a series of data mining process, the result and suggestion will return to the customers for maintenance scheduling.

In recent years, with the increasing demand for machining quality and manufacturing complication, the complexity and integration of industrial equipment has been raised dramatically [Deng et al., 2015]. On one hand, an unexpected failure can result in a devastating accident and financial losses for the company owing to the interaction behaviours among industrial equipment. On the other hand, early detection and prediction of a fault can prevent it from growing and eventually turning into critical problems [Henriquez et al., 2014]. Hence, increasing attention has been paid to condition monitoring, fault diagnosis and prognosis in modern industry [Zhao, 2014].

Simultaneously, machining centers have grown rapidly in automotive, aerospace, die making and other industries in recent years [Movahhedy and Mosaddegh, 2006]. As one of the most significant and active research fields in knowledge discovery in databases (KDD) over the last few decades, data mining and related techniques have been widely researched and applied for fault diagnosis and prognosis in machining centers.

It is well known that machine faults can result in consequences that may range from a simple replacement of a cheap bearing to an accident that will cost millions in lost production, injuries or pollution [Affonso, 2013]. It may also bother maintenance engineers to capture the trade-off between improving the system reliability and reducing the total maintenance cost simultaneously.

Accordingly, significant attention has been paid to condition-based maintenance in literature during the last few decades, and to predictive maintenance more recently [Van Horenbeek and Pintelon, 2013]. The goal of predictive maintenance is to reduce the downtime and cost of maintenance under the premise of zero failure manufacturing through monitoring the working condition of equipment and predicting when equipment failure might occur. The prediction of a future potential fault enables the planning of maintenance before the fault happens [Li et al., 2016].

Considerable progress has been made in fault interpretation, detection, and prediction for machining centers based on DM during the last few decades, especially in specific core components or performance, such as gearbox [C. Wang et al., 2012], thermal error [C.-W. Wu et al., 2012], and rolling element bearings [Abbasion et al., 2007]. However, most of these studies only focused on their own parts or concentration. There is still a lack of systematic research to guide the implementation of predictive maintenance under the Industry 4.0 era for machining centers. Moreover, it is difficult to realize all the advantages of predictive maintenance without the foundation of correlation techniques such as big data analysis and cloud-computing. Many manufacturing systems are still not ready to manage big data owing to the high demands on the access and quality of data. Furthermore, the extraction of relevant information from multiple data sources in machining tools still remains a challenge in many situations [Lee et al., 2014]. Based on these consideration, this chapter is conceived with the objective of offering effective guidelines to select suitable fault analysis techniques and implement predictive maintenance in machining centers.

4.2 Fault analysis techniques in machining centers

The term "machining center" can be used to describe any computer numerical control (CNC) milling and drilling machine that includes an automatic tool changer and a table that clamps the workpiece in place. According to the orientation of the spindles, they can be divided into two types: vertical and horizontal. Vertical machining centers generally have good precision whereas horizontal machining centers favour production. The spindle of a vertical machining center is vertically oriented. Generally, a vertical machining center includes several sub-systems that should be monitored. As shown in Figure 4.2, it may include a server motor system, ball screw system, guide systems, spindle system, tool magazine, hydraulic system, lubrication system, and cooling system [Duro et al., 2016; Shi et al., 2015]. All these systems have unique functions, and failures occurring at any one of them may cause faults in the entire machining center.

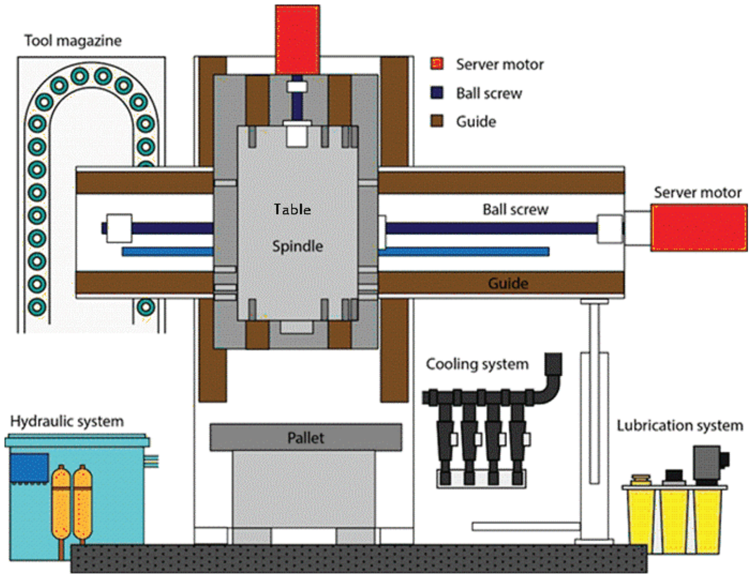

Figure 4.2 Components of a machining center

In order to achieve fault diagnosis and prognosis in machining centers, many contributions have been made in this area. Normally, research in this field can be divided into seven groups: geometric measurement analysis, vibration analysis, oil analysis, cutting fluid analysis, energy consumption analysis, temperature analysis, and acoustic emission analysis, according to the observed components, monitoring method, or specific purpose [Bort et al., 2016; Duro et al., 2016; Fan et al., 2015; K. Liu et al., 2016; Shi et al., 2015; Sparham et al., 2016; Usop et al., 2015]. However, the integration of all these techniques and information to form comprehensive, high-efficiency, and intelligent maintenance strategies still remains a challenge till the breakthrough of Industry 4.0, which combines the strengths of optimized industrial manufacturing with internet technologies and changes the manufacturing process, maintenance management, and maintenance strategies significantly. The following sections will give a brief introduction about those research objectives, which could be selected as targets or measures during the implantation of predictive maintenance.

Geometric measurement analysis

Geometrical accuracy present the capability of position for a machining center and affect the overall machining precision of machining objects directly [Schwenke et al., 2008]. Normally, a machining center may have several geometric measurement systems to monitor and measure the geometric positions of important transmission parts, or use one multi-dimensions measuring system to simultaneously measure several degrees of freedom [Khan and Chen, 2011]. Geometric measurement analysis detects, and predicts machine wear by monitoring the difference between the measurement systems, e.g. transmission error can be evaluated according to the change of difference between the measurement systems of screw and table, then compensations can be made to achieve higher precise. In addition, a systematic geometric error correction and compensation in machine tools is important to enhance the manufacturing accuracy [Zhong et al., 2015]. Among this, backlash compensation has been studied for years and solved in various methods, e.g. neural network models for backlash compensation in a gear system [Menon and Krishnamurthy, 1999]. A dynamic fuzzy logic-based adaptive algorithm was researched for backlash compensation [Suraneni et al., 2005]. Some research about backlash error prediction was also made to realize predictive maintenance in machining centers [K. Wang et al., 2015]. Other improved geometric error measuring methods were also researched and able to identify the machine tool error and make the compensation [Zhang et al., 2013].

Vibration analysis (VA)

Vibration analysis is the most well-known technology for rotating equipment maintenance. It is the most efficient technology for early prediction and detection of failures in mechanical equipment [Saimurugan et al., 2011]. Vibration analysis is commonly applied to such machining center components as shafts, bearings, and gearbox. Applied sensor technology can be selected by considering the frequency range and operating conditions. Position transducers, velocity sensors, acceleration, and spectral emission energy sensors are used for low-, Middle-, high-, and very high- frequency ranges, respectively. In addition, many new type sensors have been developed for a convenient vibration signal acquisition [Deraemaeker et al., 2010; Freundlich and Pietrzakowski, 2011; Kageyama et al., 2005]. Wireless sensors are also involved in these research fields [Aydın et al., 2015; Bocca et al., 2011]. Additionally, many CI algorithms are used for vibration analysis, e.g.

SVM [G. Wang et al., 2014], FLS [Kothamasu and Huang, 2007], neural networks [C. Liu et al., 2009] or hybrid algorithms [Jossa et al., 2000].

Oil analysis (OA)

Machine lubrication systems are a very important portion of manufacturing and production workshop maintenance. Automatic lubrication systems eliminate the need for frequent manual lubrication inspection, providing a safer, more frequent, and opportune monitored approach to machine lubrication. Oil debris monitoring can be used for the early detection and tracking of damage in bearing and gear components in machine tools. Indeed, 80% of gear box problems can be attributed to the bearings, which subsequently lead to damage to the gearing [Dupuis, 2010]. Oil monitoring constitutes an important and essential component of condition monitoring technologies and has distinguished advantages in revealing wear, lubrication and friction conditions of tribe-pairs [T. Wu et al., 2013]. Lubrication oil analysis is achieved by selecting proper sensors [Halme et al., 2010] to process the temperature signals [Sparham et al., 2014], water contamination analysis or chemical properties analysis [T. Wu et al., 2013].

In most cases, oil is pumped through the component in a close-loop system, and metal debris from creaked gearbox wheels or bearings is caught by a filer. The amount and type of metal debris can indicate the health of the component. OA has three main purpose: (1) to monitor the lubricant; condition and reveal whether the system fluid is healthy and fit for further service or requires a change; (2) to ensure the oil quality (e.g. contamination by parts, moisture); (3) to safeguard the components involved (part characterization).

Cutting fluid analysis

The primary functions a cutting fluid include the cooling and lubrication both of the workpiece and cutting tool's edge, an improvement of machined surface quality and an increase in tool life, and further a reduction in spindle power in many machining processes, which offers considerable savings when this reduction in electrical demand is accrued per annum [Smith, 2008]. Most common cutting fluid tests include:

- Concentration,
- PH (Alkalinity),

- Corrosion protection,
- Fluid stability,
- Bacteria count.

Energy consumption analysis

Machining processes consume large amounts of energy. Assessing energy consumption of machining centers is useful for evaluating the energy efficiency of various machine tools and increasing interests in green manufacturing [Behrendt et al., 2012]. Therblig-based energy demand modelling methodology can be used to evaluate energy consumption of machining processes quantitatively [Lv et al., 2014]. It is also a feasible solution to monitor the machine tool condition by monitoring the differential electrical power consumption [Al-Sulaiman et al., 2005]. Power consumption detection can be also applied to machine spindle motor for fault diagnosis [Reñones et al., 2010].

Temperature analysis (TA)

Temperature Analysis (TA) to the observed component is one of the most common method to check the requirement of maintenance. TA aids in detecting the presence of any potential failure related to temperature changes in the equipment. In machine tools, TA is applied on such components as bearings, cooling fluids, lubricating oil, motors, moving beds, fixtures, and optical pyrometers.

TA is reliable because every piece of equipment has limited operation temperature and easy to be executed. However, temperature develops slowly and is not sufficient for early and precise fault detection. Additionally, the measured temperature can also be influenced by the surroundings. Therefore, TA is rarely used alone but often as a secondary source of information. In this case, the primary source could be vibration monitoring.

Acoustic emission analysis

Acoustic Emission (AE) phenomena are based on the release of energy in the form of transitory elastic waves within a material via a dynamic deformation process. Typically, sources of AE within a material are creak initiation and propagation, breaking of fibers, and matrix creaking and fretting between surfaces at de-bonds or de-laminations. Unlike VA, AE can detect failures characterized by high-frequency vibrations range from 50 k HZ to I MHZ. Piezoelectric transducers and optic fiber displacement sensors are often employed in this approach. The most commonly measured AE parameters for diagnosis are amplitude, root means square value, energy, kurtosis, crest factor, counts and events.

This method is typically applied for fault detection in components such as gearboxes, bearings, spindles, tools and ball screw. Its advantages include a large frequency range and relatively high signal-to-noise ratio. The main limitation of AE is its cost. Furthermore, only a few types of faults occur in the high-frequency range. Another limitation of AE is the attenuation of the signal during propagation. Therefore, an AE sensor must be located as close to its source as possible, which may pose a practical constraint in applying AE to certain machines.

Some experimental studies have been researched to compare the diagnostic and prognostic capabilities of AE, VA and spectrometric OA on spur gears. It is observed that based on the analysis of root means square levels, only the AE techniques was more sensitive in detecting and monitoring faults than either the vibration or spectrometric OA [Tan et al., 2007].

4.3 Steps to implement predictive maintenance in machining centers

As discussed above, monitoring systems in machining centers may require DM methods for fault diagnosis and prognosis according to different monitoring purposes or components. DM and CI could be applied to discover failures information and optimize the solutions respectively. Here, the key steps to implement predictive maintenance in machining centers based on Industry 4.0 concepts is formulated as following to provide guidelines for researchers and practitioners :

- Sensor selection and data acquisition
- Data preprocessing
- Data mining
- Data record and publication
- Decision support
- Maintenance implementation

4.3.1 Sensor selection and data acquisition

This is the first step to implement diagnosis and prognosis based on DM in machining centers. The task in this step is to select a suitable sensor and optimal data collection strategy to extend the physical world using a variety of sensing, detection, and identification techniques, and connect the objects or enable them to interact with each other. The data acquisition process transforms the sensor signals into domains that have the most information to represent the condition of the equipment or a fusion of several domains. Various sensors such as micro-sensors, ultrasonic sensors, vibration sensors, and acoustic emission sensors can be designed to collect different data. The selection of sensors determines the representation of the machine health by the collected data, considering both the specifications and cost-effectiveness. Moreover, with the increase in the complexity of machine systems, the sensor network is considered as a feasible solution for condition monitoring in machining centers, which may include different kinds of sensors. Sensor fusion achieves significance under this condition. Therefore, obtaining smart sensors is also a prominent research field in condition monitoring [Son et al., 2009]. The selection of suitable sensors and data sources is significant to the effectiveness of condition monitoring, and a complete data acquisition system could directly improve the correction and efficiency of diagnosis and prognosis.

4.3.2 Data preprocessing

After the data acquisition, all the collected data will be stored in the data warehouse for diagnosis and prognosis. However, during the process of knowledge discovery, if there is too much irrelevant and redundant information, such as noise or unreliable data, the training phase will be more challenging. Therefore, it is necessary to preprocess the data before the subsequent step. Generally, the major functions involved in data preprocessing include data cleaning, data integration, data reduction, and data transformation. The development of storage media and computation ability results in massive data during the data acquisition process. Data preprocessing can effectively clean the raw data, reduce the dimension of the data, and store it back in the warehouse for knowledge discovery. Therefore, massive data can be converted to features or statistical values as the input variables of the DM process.

4.3.3 Data mining

DM has the capability to discover hidden links, recognize unknown patterns, and predict future trends by digging through and analyzing enormous sets of data [Sumathi and Sivanandam, 2006]. The functions, or models, of DM can be categorized according to the task performed [Siguenza-Guzman et al., 2015], such as clustering, classification, decision trees, predication, regression, and association. More detail and systematic knowledge about the application of data mining in predictive maintenance have been discussed in Chapter 3.

In this step, data mining mainly focus on the detection, identification and prediction for potential or impending failures. Fault diagnosis and prognosis strategies have been developed and found extensive utility in a wide range of application domains in recent years. Model-based technique could take advantage of the actual system and model in machining centers to generate the difference between the two outputs, which is indicative of a potential fault condition. When the high-accuracy dynamic system model is unavailable or difficult to establish, data-driven or hybrid techniques could also be leveraged to map anticipated fault conditions. Since the historical data could be obtained easily in a machining center, the data-driven models could be very useful to identify impending faults and evaluate working conditions for machining centers.

Some common diagnosis and prognosis algorithms are listed in Chapter 2. All these techniques have been widely applied or already demonstrated their ability to deal with certain issues. In Chapter 5, a case study about DBN-based backlasher error prediction in a vertical machining center will be demonstrated and prove the superiority of deep learning architecture in predictive maintenance. In Chapter 6, DNN will also be leveraged to identify fault types and recognize fault severity ranking in a rotating equipment along with LSTM-based anomaly detection.

4.3.4 Data record and publication

After data mining, the information of potential failures could be recorded and published on online dashboards for remote condition monitoring and information sharing. As shown in Figure 4.3, the predicted backlash errors in a machining center are published on the dashboard to share the result of fault prediction. Management systems in different levels such as enterprise resource planning and manufacturing execution system could reschedule their production plan based on the failure information. In addition, local RFID readers can collect and identify all the RFID tags, which represent the main target components in the equipment as shown in Figure 4.4. Figure 4.5 shows a sample of Information in a RFID Tag. Here, Tag ID "0xE20093747411026111005007" can be used for components identity. Normally, the 512 bits user data are reserved for the customized information, which can be converted to 64 ASCII characters. In this case, the condition evaluation related information is written in the user memory of the RFID tag. The information comprises the current working condition, whether has potential faults in the prediction period, when the potential faults may happen, evaluation time and date, in the form of "COND=0 POT=1 FAULT=28.05.16 TIME=14:35 DATE=21.05.16". Since the tag memory is written in the form of hexadecimal, the texts have to be converted to HEX before writing. In this case, the current working condition of the component, which is a spindle in a machining center, is normal (COND=0). Potential faults are detected (POT=1), which may happen in 28.05.2016 (FAULT=28.05.16). In addition, the detection is taken at 14:35 on 21.05.2016 (TIME=14:35 DATE=21.05.16).

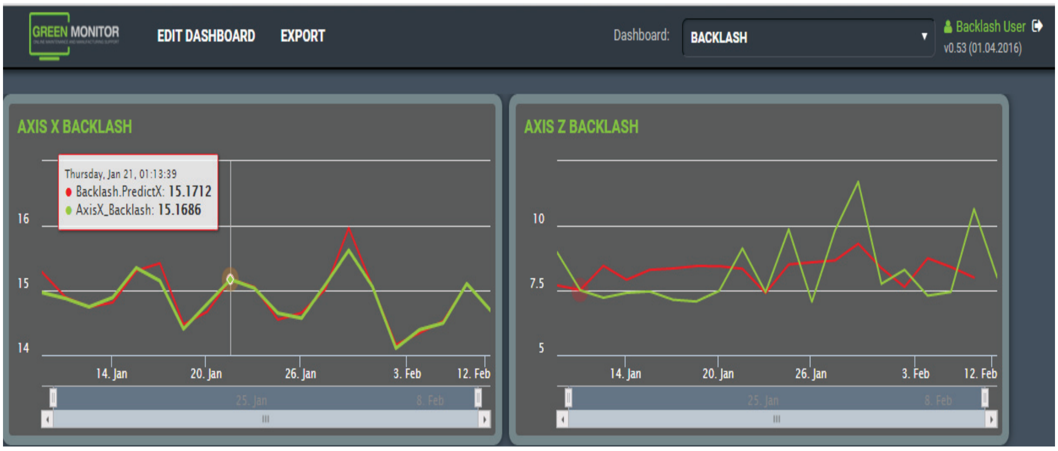

Figure 4.3 Monitoring dashboard

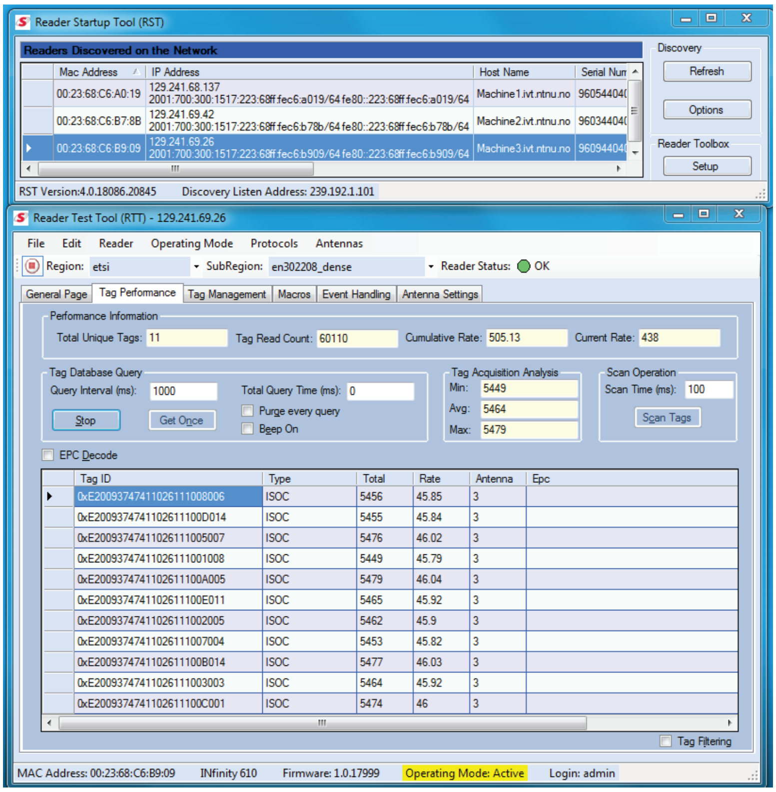

Figure 4.4 Record of tags in RFID readers

Combining the result of diagnosis and prognosis with logistics information collected from RFID tags, customers can figure out when, where, which equipment, and which components may have faults. The recommended maintenance strategy could be formed in the decision support system according to the prediction or evaluation result.

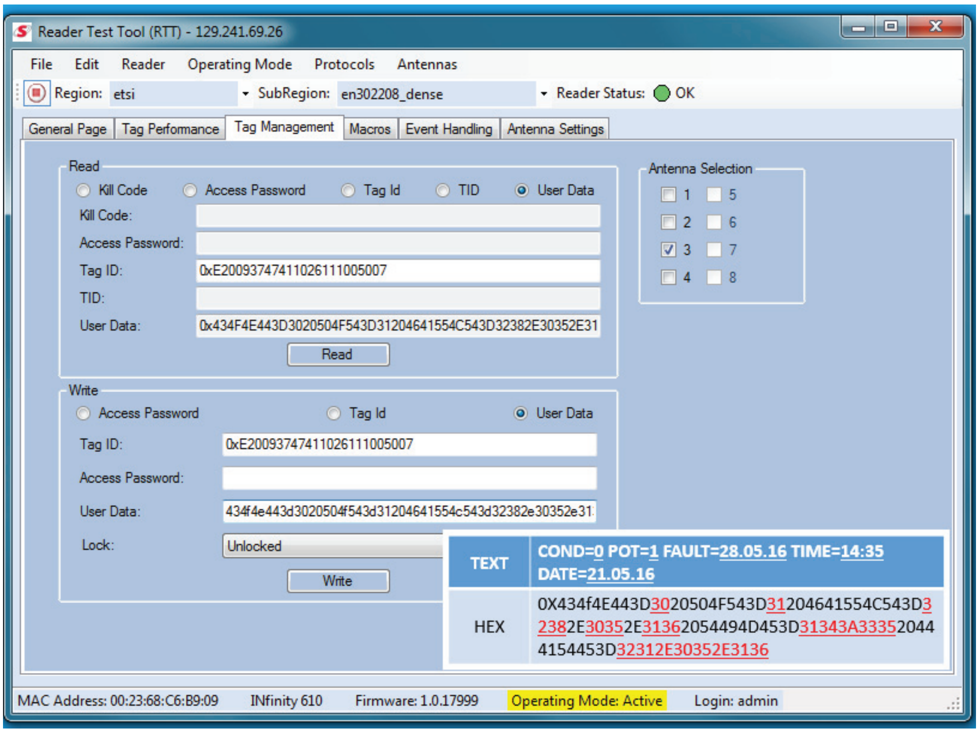

Figure 4.5 Information in a RFID tag

4.3.5 Decision support

The main target in this step is to visualize the result of DM and provide an optimized maintenance strategy for machining centers according to the result of DM. It can also be considered as the application of IoS. Generally, a diagram of key performance indicator (KPI), also called a spider chart, can be used for presenting the situation of equipment. The conditions of equipment can be defined in several levels from zero to one. For example, zero indicates no faults and one indicates complete damage of equipment. The KPI may be formed according to the outputs of the DM. The diagram will enable operators or managers to evaluate the performance visually, and subsequently, an optimized maintenance schedule can be provided according to the result of evaluation.

Maintenance planning and scheduling optimization is a kind of nondeterministic polynomial time (NP) problem and it is always difficult for the decision-makers to capture the trade-off. SI algorithms could be a very good technique to solve this kind of problem. Usually, one may apply genetic algorithm, particle swarm optimization, ant colony optimization, and bee colony algorithm as decision support methods, and attempt to determine the optimal dynamic predictive maintenance scheduling. All these methods are selectable to solve maintenance scheduling optimization problems. Furthermore, this step may also include the function of failure identification and the evaluation of performance degradation according to the result of DM. In Chapter 5, a novel HDPS-BPSO maintenance scheduling approach will be proposed and leveraged in practical application to demonstrate the advantage of predictive maintenance through the comparison with preventive one.

4.3.6 Maintenance implementation

In this step, maintenance will be implemented after the decision-makers choose the strategy of maintenance. It can be considered as the purpose of CPS. The physical world is transferred into the virtual one for communication, computation, analysis, and decision-making via the previous steps. In this step, we react to the physical world according to the result of previous steps and implement maintenance to achieve a certain purpose, e.g., to minimize the cost of maintenance, realize zero-defect manufacturing, or reduce breakdown.

Moreover, this step may also include the function of error correction, compensation, and feedback control based on the results from the maintenance decision support to continue to run the machining center and process in a normal condition. Some techniques can also be used to correct and compensate certain errors, e.g. artificial neural network can be used for the compensation of geometric errors in computer-controlled machining centers. However, the error correction and compensation process is mainly dependent on the types of machines and processes, so this step should also take control devices and the maintenance management system in consideration.

4.4 Summary

DM plays a very important role for predictive maintenance in machining centers owing to their complexity and high machining precision. The theoretical contribution of this chapter could be represented by the collection, classification, and induction of DM approaches applied for fault identification and prediction in machine centers. This chapter is conceived with the target of offering effective guidelines to formulate systematic fault diagnosis and prognosis steps to implement predictive maintenance in machining centers based on DM result. The guidelines are coincident with the general framework proposed in Chapter 3 but focuses more on the diagnosis and prognosis stage in machining centers. In next chapter, a hierarchical diagnosis and prognosis system for backlash error detection and prediction in machining centers based on deep learning will be introduced in a practical case study along with a novel HDPS-BPSO maintenance scheduling strategy for predictive maintenance implementation.

References

Abbasion, S., Rafsanjani, A., Farshidianfar, A., and Irani, N. (2007). Rolling element bearings multi-fault classification based on the wavelet denoising and support vector machine. *Mechanical Systems and Signal Processing, 21*(7), 2933-2945.

Affonso, L. O. A. (2013). *Machinery Failure Analysis Handbook: sustain your operations and maximize uptime*: Elsevier.

Al-Sulaiman, F. A., Baseer, M. A., and Sheikh, A. K. (2005). Use of electrical power for online monitoring of tool condition. *Journal of Materials Processing Technology, 166*(3), 364-371.

Aydın, İ., Karaköse, M., and Akın, E. (2015). Combined intelligent methods based on wireless sensor networks for condition monitoring and fault diagnosis. *Journal of Intelligent Manufacturing, 26*(4), 717-729.

Behrendt, T., Zein, A., and Min, S. (2012). Development of an energy consumption monitoring procedure for machine tools. *CIRP Annals-Manufacturing Technology, 61*(1), 43-46.

Bocca, M., Eriksson, L. M., Mahmood, A., Jäntti, R., and Kullaa, J. (2011). A synchronized wireless sensor network for experimental modal analysis in structural health monitoring. *Computer‐Aided Civil and Infrastructure Engineering, 26*(7), 483-499.

Bort, C. M. G., Leonesio, M., and Bosetti, P. (2016). A model-based adaptive controller for chatter mitigation and productivity enhancement in CNC milling machines. *Robotics and Computer-Integrated Manufacturing, 40*, 34-43.

Deng, P., Ren, G., Yuan, W., Chen, F., and Hua, Q. (2015). An integrated framework of formal methods for interaction behaviors among industrial equipments. *Microprocessors and Microsystems, 39*(8), 1296-1304.

Deraemaeker, A., Preumont, A., Reynders, E., De Roeck, G., Kullaa, J., Lamsa, V., . . . Papatheou, E. (2010). Vibration-based structural health monitoring using large sensor networks. *Smart Structures and Systems, 6*(3), 335-347.

Dupuis, R. (2010). *Application of oil debris monitoring for wind turbine gearbox prognostics and health management.* Paper presented at the Proceedings of the annual conference of the prognostics and health management society.

Duro, J. A., Padget, J. A., Bowen, C. R., Kim, H. A., and Nassehi, A. (2016). Multi-sensor data fusion framework for CNC machining monitoring. *Mechanical Systems and Signal Processing, 66*, 505-520.

Fan, K., Yang, J., and Yang, L. (2015). Unified error model based spatial error compensation for four types of CNC machining center: Part I—Singular function based unified error model. *Mechanical Systems and Signal Processing, 60*, 656-667.

Freundlich, J., and Pietrzakowski, M. (2011). Monitoring of Roof Truss Girder Vibration Using Piezoelectric Sensors–FEM Simulation *Vibration Problems ICOVP 2011* (pp. 441-446): Springer.

Halme, J., Gorritxategi, E., and Bellew, J. (2010). Lubricating Oil Sensors *E-maintenance* (pp. 173-196): Springer.

Henriquez, P., Alonso, J. B., Ferrer, M. A., and Travieso, C. M. (2014). Review of automatic fault diagnosis systems using audio and vibration signals. *Systems, Man, and Cybernetics: Systems, IEEE Transactions on, 44*(5), 642-652.

Jossa, I., Marschner, U., and Fischer, W.-J. (2000). Application of the FSOM to machine vibration monitoring *Fuzzy Control* (pp. 397-405): Springer.

Kageyama, K., Murayama, H., and Uzawa, K. (2005). A New Fiber-Optic Acoustic/Vibration Sensor—Characteristics and Application to Civil Structural Health Monitoring *Sensing Issues in Civil Structural Health Monitoring* (pp. 443-

452): Springer.

Khan, A. W., and Chen, W. (2011). A methodology for systematic geometric error compensation in five-axis machine tools. *The International Journal of Advanced Manufacturing Technology, 53*(5-8), 615-628.

Kothamasu, R., and Huang, S. H. (2007). Adaptive Mamdani fuzzy model for condition-based maintenance. *Fuzzy Sets and Systems, 158*(24), 2715-2733.

Lee, J., Kao, H.-A., and Yang, S. (2014). Service innovation and smart analytics for industry 4.0 and big data environment. *Procedia CIRP, 16*, 3-8.

Li, Z., Wang, K., and He, Y. (2016). Industry 4.0–Potentials for Predictive Maintenance. *International Workshop of Advanced Manufacturing and Automation*.

Liu, C., Jiang, D., and Zhao, M. (2009). *Application of RBF and Elman neural networks on condition prediction in CBM.* Paper presented at the The Sixth International Symposium on Neural Networks (ISNN 2009).

Liu, K., Sun, M., Zhu, T., Wu, Y., and Liu, Y. (2016). Modeling and compensation for spindle's radial thermal drift error on a vertical machining center. *International Journal of Machine Tools and Manufacture, 105*, 58-67.

Lv, J., Tang, R., and Jia, S. (2014). Therblig-based energy supply modeling of computer numerical control machine tools. *Journal of Cleaner Production, 65*, 168-177.

Menon, K., and Krishnamurthy, K. (1999). Control of low velocity friction and gear backlash in a machine tool feed drive system. *Mechatronics, 9*(1), 33-52.

Movahhedy, M. R., and Mosaddegh, P. (2006). Prediction of chatter in high speed milling including gyroscopic effects. *International Journal of Machine Tools and Manufacture, 46*(9), 996-1001.

Reñones, A., Rodríguez, J., and de Miguel, L. J. (2010). Industrial application of a multitooth tool breakage detection system using spindle motor electrical power consumption. *The International Journal of Advanced Manufacturing Technology, 46*(5-8), 517-528.

Saimurugan, M., Ramachandran, K., Sugumaran, V., and Sakthivel, N. (2011). Multi component fault diagnosis of rotational mechanical system based on decision tree and support vector machine. *Expert Systems with Applications, 38*(4), 3819-3826.

Schwenke, H., Knapp, W., Haitjema, H., Weckenmann, A., Schmitt, R., and Delbressine, F. (2008). Geometric error measurement and compensation of machines—an update. *CIRP Annals-Manufacturing Technology, 57*(2), 660-675.

Shi, S., Lin, J., Wang, X., and Xu, X. (2015). Analysis of the transient backlash error in CNC machine tools with closed loops. *International Journal of Machine Tools and Manufacture, 93*, 49-60.

Siguenza-Guzman, L., Saquicela, V., Avila-Ordóñez, E., Vandewalle, J., and Cattrysse, D. (2015). Literature review of data mining applications in academic libraries. *The Journal of Academic Librarianship, 41*(4), 499-510.

Smith, G. T. (2008). *Cutting tool technology: industrial handbook*: Springer Science & Business Media.

Son, J.-D., Niu, G., Yang, B.-S., Hwang, D.-H., and Kang, D.-S. (2009). Development of smart sensors system for machine fault diagnosis. *Expert systems with applications, 36*(9), 11981-11991.

Sparham, M., Sarhan, A. A., Mardi, N., Dahari, M., and Hamdi, M. (2016). Cutting force analysis to estimate the friction force in linear guideways of CNC machine. *Measurement, 85*, 65-79.

Sparham, M., Sarhan, A. A., Mardi, N., and Hamdi, M. (2014). Designing and manufacturing an automated lubrication control system in CNC machine tool guideways for more precise machining and less oil consumption. *The International Journal of Advanced Manufacturing Technology, 70*(5-8), 1081-1090.

Sumathi, S., and Sivanandam, S. (2006). *Introduction to data mining and its applications* (Vol. 29): Springer.

Suraneni, S., Kar, I. N., Murthy, O. R., and Bhatt, R. (2005). Adaptive stick–slip friction and backlash compensation using dynamic fuzzy logic system. *Applied Soft Computing, 6*(1), 26-37.

Tan, C. K., Irving, P., and Mba, D. (2007). A comparative experimental study on the diagnostic and prognostic capabilities of acoustics emission, vibration and spectrometric oil analysis for spur gears. *Mechanical Systems and Signal Processing, 21*(1), 208-233.

Usop, Z., Sarhan, A. A., Mardi, N., and Wahab, M. N. A. (2015). Measuring of positioning, circularity and static errors of a CNC Vertical Machining Centre for validating the machining accuracy. *Measurement, 61*, 39-50.

Van Horenbeek, A., and Pintelon, L. (2013). A dynamic predictive maintenance policy for complex multi-component systems. *Reliability Engineering & System Safety, 120*, 39-50.

Wang, C., Kang, Y., and Chung, Y. (2012). *The Gearbox Fault Detection of Machine Center by Using Time Frequency Order Spectrum.* Paper presented at the Advanced Materials Research.

Wang, G., Yang, Y., Zhang, Y., and Xie, Q. (2014). Vibration sensor based tool condition monitoring using ν support vector machine and locality preserving projection. *Sensors and Actuators A: Physical, 209*, 24-32.

Wang, K., Li, Z., Braaten, J., and Yu, Q. (2015). Interpretation and compensation of backlash error data in machine centers for intelligent predictive maintenance using ANNs. *Advances in Manufacturing, 3*(2), 97-104.

Wu, C.-W., Tang, C.-H., Chang, C.-F., and Shiao, Y.-S. (2012). Thermal error compensation method for machine center. *The International Journal of Advanced Manufacturing Technology, 59*(5-8), 681-689.

Wu, T., Wu, H., Du, Y., and Peng, Z. (2013). Progress and trend of sensor technology for on-line oil monitoring. *Science China Technological Sciences, 56*(12), 2914-2926.

Zhang, Y., Yang, J., and Zhang, K. (2013). Geometric error measurement and compensation for the rotary table of five-axis machine tool with double ballbar. *The International Journal of Advanced Manufacturing Technology, 65*(1-4), 275-281.

Zhao, C. (2014). *Fault subspace selection and analysis of relative changes based reconstruction modeling for multi-fault diagnosis.* Paper presented at the 26th

Chinese Control and Decision Conference (2014 CCDC).

Zhong, G., Wang, C., Yang, S., Zheng, E., and Ge, Y. (2015). Position geometric error modeling, identification and compensation for large 5-axis machining center prototype. *International Journal of Machine Tools and Manufacture, 89*, 142-150.

Chapter 5
Implementation of predictive maintenance for backlash error

This chapter demonstrates a case study of applying DBN for backlash error prediction in a machining center along with a novel HDPS-BPSO maintenance scheduling strategy for predictive maintenance implementation. It proves the superiority of the deep learning model for fault forecast, when important parameters are missing and the target condition is beyond training data. The case study is partly retrieved from [Z. Li et al., 2017].

5.1 Background

Geometric errors that occur in machining centers are the errors on account of the inaccuracies built in during assembly and from the components used in the machine. The errors may be affected by many error sources [Lee and Yang, 2013; Zhu et al., 2012]. These error sources may cause a series of changes in the geometry of the components and present in the structural loop, including the spindle shaft, the ball screws, the bearings, the housing, the guideways and frame, and work-holding fixtures. The errors may not only cause significant quality and accuracy degradation but also fatal breakdown of machines, which can lead to serious economic loss [Huang et al., 2015; Jiang and Cripps, 2015; Mourtzis et al., 2016]. Therefore, it is especially crucial to accurately detect the existence of geometric errors as early as possible and predict the error in a period of working time [Cheng et al., 2014; Siguenza-Guzman et al., 2015; Zhong et al., 2015]. Schwenke et al. [2008] reviewed various technologies to evaluate the geometric errors of machines and their basic characteristics. As reported in that paper, backlash usually affect uncertainties in the measured parameters since they are usually not modelled adequately. The uncertainties may cause correlations or just erroneous estimations. Therefore, many companies choose the preventive maintenance for backlash error in machining centers, which means the error would be checked and eliminated from time to time following planned guidelines. However, this strategy is both costly and time-consuming.

During the last few decades, many researchers have studied methods to monitor, model and control backlash error for mechanical systems, and many diagnosis approaches have been proposed [Chen et al., 2016; Fines and Agah, 2008; Huanlao Liu et al., 2010; Prasanga et al., 2013; Slamani et al., 2012]. Prasanga et al. [2013] proposed a method to compensate the backlash error through two parallel thrust wires without any encoder or force sensor at the end effector. Slamani et al. [2012] evaluated the backlash error of an industrial serial robot under various conditions using a laser interferometer measurement instrument and represent the relationship between the backlash error and the robot configuration with a polynomial model. All these methods require additional measurement equipment or system along with elaborate engineering and considerable domain expertise.

Nowadays, with the trend of smart manufacturing, companies are increasingly using sensors and wireless technologies to capture data at all stages of a product's life [Kusiak, 2017]. For this reason, "Big Data" has attracted not only researchers' but also manufacturers' attention along with the development of data-driven methods from various perspectives such as product lifecycle management [J. Li et al., 2015], manufacturing [Tao et al., 2017], and maintenance [Mosallam et al., 2016]. Some machine learning and AI approaches, such as neural networks, SVM and FLS, also have been applied in backlash error evaluation or prediction. Chen et al. [2016] compensated the backlash nonlinearity by a smooth backlash inverse with the help of parameter estimations and fuzzy logic system-based approximation for an active vibration isolation system. Fines and Agah [2008] applied artificial neural network for positioning error compensation in a machine tool, and proved the feasibility to calculate compensation values. Liu et al. [2010] employed back-propagation neural network to map the backlash error in a vertical machining center and compared the result with polynomial models.

However, most of these researches only focus on the diagnosis or evaluation of current or historical backlash error. It still lacks a method with high generalization for backlash error prediction, especially when the target condition is beyond the historic data. Therefore, the method leveraged here will focus on the detection of both current and future geometric error for backlash error compensation and maintenance in machining centers.

5.2 Backlash error in machining centers

In order to perform an error mapping and subsequent compensation for backlash error, an understanding of the sources and effects of backlash error in machining centers are necessary. In mechanical engineering, backlash is a kind of nonlinear position dependent-error caused by the existence of clearance between two mechanical elements. It may occur in a rotational mechanical element as well as in a translational mechanical element [Kao et al., 1996]. Normally, a machining center is equipped with various high-precision sensors such as gratings, rotary encoders, current sensors, temperature sensors, and linear scales to guarantee the accuracy through closed or half closed loop control. Many parameters such as machine temperature, geometric position of ball screw, torque and current, can be obtained directly from the control system. In machining centers, backlash error can be acquired through some extra methods, most of which are either time-costly such as laser interferometer or only yield the maximum value [Huanlao Liu et al., 2010].

In a machining center, backlash error occurs when there exists a gap between the ball screw and spindle at the kinematic pair. As shown in Figure 5.1, when the direction of motion reversed, the spindle will not move until the gap is taken up in the opposite direction. The distance that the ball screw travels before the table will move again is the geometric error caused by backlash, which is also called as backlash error. In general, all loosely connected elements in the driving mechanism may influence the backlash error of the system.

Backlash error varies at different axis positions and depends on the moving direction. In addition, the error affects the contouring accuracy and increases over time due to wear in the machining center, which means it is almost impossible to establish an accurate physical model for backlash error prediction. Therefore, it is significant and necessary to apply machine-learning approaches to monitor, model, and predict backlash error in mechanical systems to maintain the desired level of accuracy.

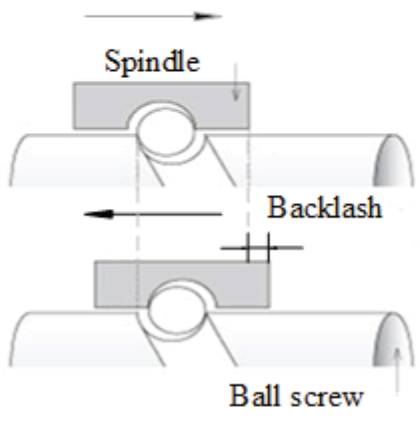

Figure 5.1 Backlash error in machining centers

5.3 Hierarchical diagnosis and prognosis system (HDPS)

As mentioned above, since it is intractable to establish physical model for backlash error, and the backlash error varies with different axis positions, most company chooses to check and eliminate backlash according to the planned guidelines, which is both costly and time-consuming. Therefore, it still lacks method to predict the backlash error instead of detecting from time to time. In this section, a novel HDPS is proposed for backlash error detection and predication in machining centers based on DBN models. The purpose of the system is to make maintenance decision based on the result of faults diagnosis and prognosis. With the help of HDPS, the maintenance team can prevent occurrence and development of failures effectively, ensure the safety of equipment and personnel, and reduce economic loss caused by failures. It can leverage fault diagnosis, performance assessment of degrading level, fault prognosis models to reach near-zero-breakdown performance and improve productivity for a company.

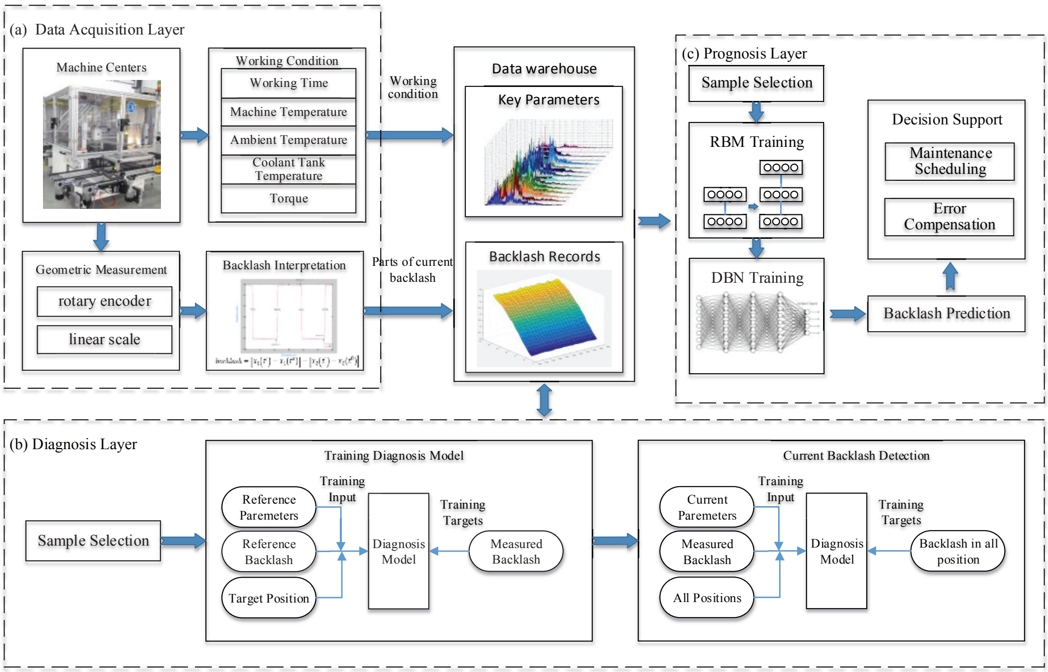

Figure 5.2 System structure of HDPS

As shown in Figure 5.2, HDPS can be divided into three layers, which are data acquisition layer, diagnosis layer and prognosis layer. In data acquisition layer, the data, including all the information require during the diagnosis and prognosis, is collected from machining centers. This is the first step to achieve diagnosis and prognosis based on data mining for machining centers. The first task of this layer is to select and collect suitable parameters which can represent the current working condition or install additional sensors for this purpose. Then, the backlash error at current axis position can be calculated form geometric measurement information through backlash error interpretation method, which will be introduced in next section. All the measured parameters and obtained backlash error are stored in the data warehouse.

Since backlash error varies at different axis positions and the measurement is both costly and time-consuming, it is almost impracticable to collect the backlash error at all positions. Therefore, in data acquisition layer, only one or several axis positions' backlash error may be acquired, and the diagnosis layer is responsible to fill up the others. At first, training samples require to be selected from the data warehouse and divided into groups for training and testing the diagnosis model. Several data-driven models like BPNN, SVMR and DBN, can be applied as diagnosis model here according to the user. Once the diagnosis model is trained by historical data, it can be used to detect current backlash error in all position and fill up all missing backlash errors back to the data warehouse.

In prognosis layer, the historical data will be selected from the data warehouse to train the prognosis model. Then the data can be used to pre-train the RBM in an unsupervised method. The DBN model can be constructed by stacking these RBM and a final decision layer, which may adjust the weight of the network according to the target values. Once the prognosis is trained, it can be applied to predict the backlash error in the future through the current working condition. In next section, the experiment for backlash error detection and prediction during our research is introduced in detail to illustrate how the HDPS and deep learning approach can work for backlash error detection and prediction in machining centers. The numerical results, which will be detailed in following sections, conform the effectiveness and feasibility of the proposed method.

5.4 DBN-based fault diagnosis and prognosis

DBN is a deep learning structure to alleviate the problem of gradients vanishing through unsupervised pre-training for a hierarchy network, which is first proposed by Hinton in 2006 [G. E. Hinton et al., 2006]. This type of deep neural network is constructed through training and stacking several layers of RBM in a greedy manner. In stage k, the output of the previous layer is used as the input to train the k^{th} layer (with respect to an unsupervised criterion), while the previous layers are kept fixed. Once this stack of RBM is trained, it can be used to initialize a multi-layer neural network for classification or regression [Erhan et al., 2010].

5.4.1.1 Restricted Boltzmann machine

Restricted Boltzmann machine is a special type of Markov random field, which consists of two layers, one with stochastic visible or observable units and the other with stochastic hidden units [Keyvanrad and Homayounpour, 2014]. RBM can be represented as bipartite graphs as shown in Figure 5.3, where all visible units v are connected to all hidden units h, and there are no visible-visible or hidden-hidden connections [D. Yu and Deng, 2011]. w_{ij} represents the interaction term between visible unit v_i and hidden unit h_j, while vector **a** and **b** are bias terms for hidden units and visible units respectively.

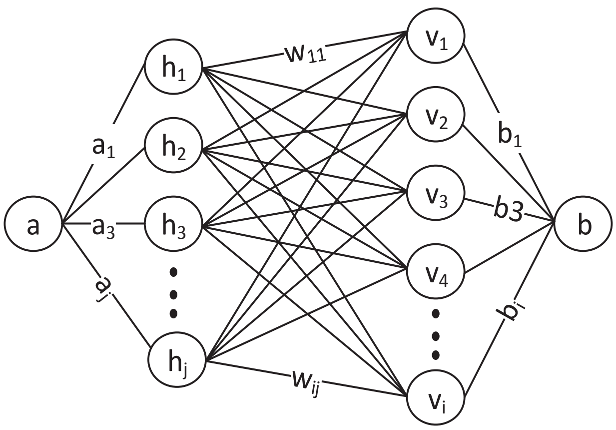

Figure 5.3 Structure of RBM

The energy of the joint configuration with bias in RBM is defined as equation (5.1):

$$E(\boldsymbol{v}, \boldsymbol{h}; \boldsymbol{W}, \boldsymbol{a}, \boldsymbol{b}) = -\boldsymbol{v}^T \boldsymbol{W} \boldsymbol{h} - \boldsymbol{b}^T \boldsymbol{v} - \boldsymbol{a}^T \boldsymbol{h}$$

$$= -\sum_{i=1}^{I}\sum_{j=1}^{J} w_{ij}v_i h_j - \sum_{i=1}^{I} b_i v_i - \sum_{j=1}^{J} a_j h_j \tag{5.1}$$

Where **W** is the concurrent weights between visible and hidden units. I and J represent the numbers of visible and hidden units. Then the joint probability distribution for all visible and hidden pairs can be defined as follows:

$$p(\boldsymbol{v}, \boldsymbol{h}; \boldsymbol{W}, \boldsymbol{a}, \boldsymbol{b}) = Z^{-1} e^{-E(\boldsymbol{v},\boldsymbol{h};\boldsymbol{W},\boldsymbol{a},\boldsymbol{b})} \tag{5.2}$$

Where Z is the partition, which can be obtained by summing all possible pairs of visible and hidden units as equation (5.3):

$$Z = \sum_{\boldsymbol{v}}\sum_{\boldsymbol{h}} e^{-E(\boldsymbol{v},\boldsymbol{h};\boldsymbol{W},\boldsymbol{a},\boldsymbol{b})} \tag{5.3}$$

Then, the probability assigned from the network for the visible vector **v** can be obtained by marginalizing out the hidden vector:

$$p(\boldsymbol{v}; \boldsymbol{W}, \boldsymbol{a}, \boldsymbol{b}) = \sum_{h} p(\boldsymbol{v}, \boldsymbol{h}; \boldsymbol{W}, \boldsymbol{a}, \boldsymbol{b}) = Z^{-1} \sum_{\boldsymbol{h}} e^{-E(\boldsymbol{v},\boldsymbol{h};\boldsymbol{W},\boldsymbol{a},\boldsymbol{b})} \tag{5,4}$$

Due to the specific structure of RBM, there are no direct connection between hidden units. Therefore, all the visible and hidden units are conditionally independent [G. Hinton, 2010], and the conditional probabilities can be efficiently calculated as equation (5.5) and (5.6):

$$p\left(h_j = 1 | \boldsymbol{v}; \boldsymbol{W}, \boldsymbol{a}, \boldsymbol{b}\right) = S\left(a_j + \sum_{i=1}^{I} v_i w_{ij}\right) \tag{5.5}$$

$$p(v_i = 1 | \boldsymbol{h}; \boldsymbol{W}, \boldsymbol{a}, \boldsymbol{b}) = S\left(b_i + \sum_{j=1}^{J} h_j w_{ij}\right) \tag{5.6}$$

Where $S(x)$ is the logistic sigmoid function $S(x) = 1/(1 + e^{-x})$

Then, the RBM model with binary units can be learned through negative log-likelihood gradients [Gan and Wang, 2016]. The derivative of the log probability of a training vector can be obtained as follows:

$$\frac{\partial \log p(\boldsymbol{v}; \boldsymbol{W}, \boldsymbol{a}, \boldsymbol{b})}{\partial w_{ij}} = \langle v_i h_j \rangle_{data} - \langle v_i h_j \rangle_{model} \tag{5.7}$$

$$\frac{\partial \log p(\boldsymbol{v}; \boldsymbol{W}, \boldsymbol{a}, \boldsymbol{b})}{\partial \boldsymbol{a}} = \langle h_j \rangle_{data} - \langle h_j \rangle_{model} \tag{5.8}$$

$$\frac{\partial \log p(\boldsymbol{v}; \boldsymbol{W}, \boldsymbol{a}, \boldsymbol{b})}{\partial \boldsymbol{b}} = \langle v_i \rangle_{data} - \langle v_i \rangle_{model} \tag{5.9}$$

Where the angle brackets denote the expectations with the distribution specified by the subscript that follows. Then the learning rule for all parameters can be obtained as follows:

$$\Delta w_{ij} = \mathcal{E}_w\big(\langle v_i h_j \rangle_{data} - \langle v_i h_j \rangle_{model}\big) \tag{5.10}$$

$$\Delta a_{ij} = \mathcal{E}_a\big(\langle h_j \rangle_{data} - \langle h_j \rangle_{model}\big) \tag{5.11}$$

$$\Delta b_{ij} = \mathcal{E}_b(\langle v_i \rangle_{data} - \langle v_i \rangle_{model}) \tag{5.12}$$

Where $\mathcal{E}_w$, $\mathcal{E}_a$ and $\mathcal{E}_b$ represent learning rate of weight, hidden bias and visible bias, respectively. According to the previously mentioned RBM property, an unbiased sample of $\langle\, .\rangle_{data}$with the respect to the data distribution can be easily obtained, while attaining an unbiased sample of $\langle\, .\rangle_{model}$ is intractable, since it can be done through starting from any random state of the visible units and performing sequential Gibbs sampling for a long time [Keyvanrad and Homayounpour, 2014]. Therefore, the Contrastive Divergence (CD) method [G. E. Hinton, 2002] is applied to approximate the gradient objective function, where $\langle\, .\rangle_{model}$ is replaced by k iterations of Gibbs sampling. During Gibbs sampling, each iteration updates all hidden units according to equation (5.11), followed by updating of all visible unites through equation (5.12), as shown in Figure 5.4. Although CD method is not a perfect gradient computation method, the results has been proved acceptable [Carreira-Perpinan and Hinton, 2005].

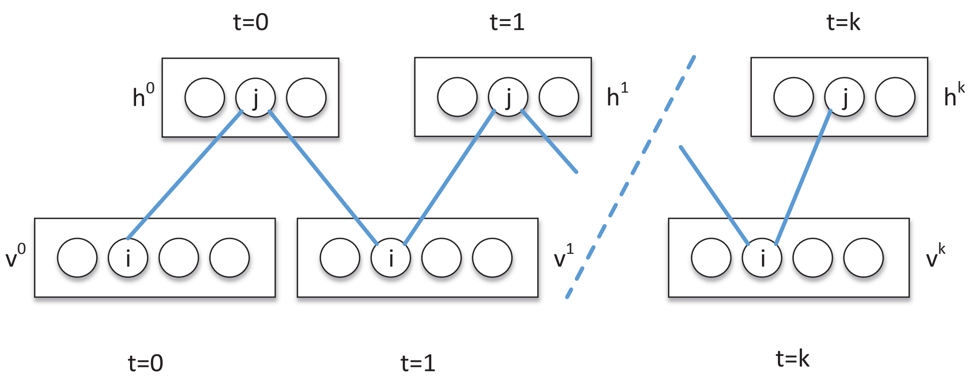

Figure 5.4 Contrastive divergence training

5.4.1.2 DBN construction for fault diagnosis and prognosis

As mentioned above, a DBN can be constructed through stacking RBM, each of which contains one visible layer and one hidden layer respectively. The construction process of DBN is well described in [G. Hinton et al., 2012]. Each RBM is pre-trained with their own training data by CD training algorithm, and its output serves as the training data for the next RBM layer. As shown in Figure 5.5, the input layer and the first hidden layer h_1 construct the first RBM. Then the states of the binary hidden units of the first trained RBM is used to train the next hidden layer h_2, then hidden layer h_1 and hidden layer h_2 form the second RBM. These layer by layer unsupervised training method can effectively pre-train the DBN.

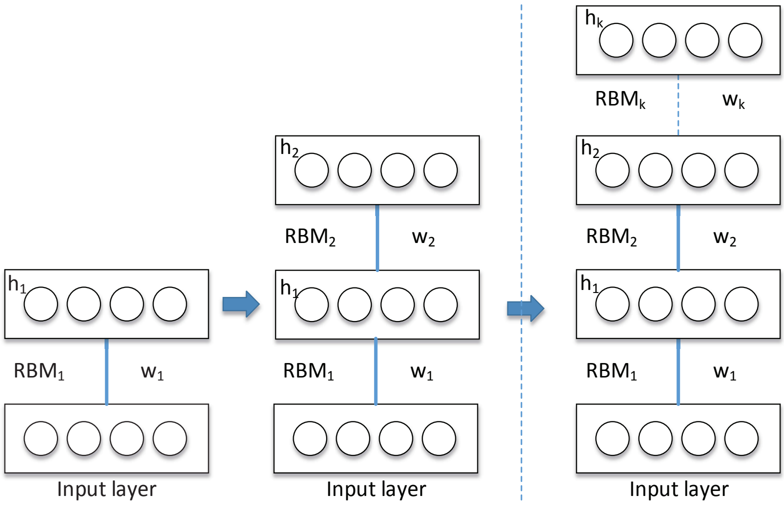

Figure 5.5 Overall construction of DBN

To achieve high classification or regression performance for fault diagnosis and prognosis, a final decision layer with variables, which represent the desired outputs or labels, is added to the stacked RBM. The final structure of DBN for fault diagnosis and prognosis in shown in Figure 5.6, which is composed of several successive RBM layers and a final decision layer for faults clustering, faulty component's identity, or evaluation of potential failures. Once the DBN is initialized, the BP algorithm, which is a supervised learning method and applied in BPNN, can be employed to adjust the weights.

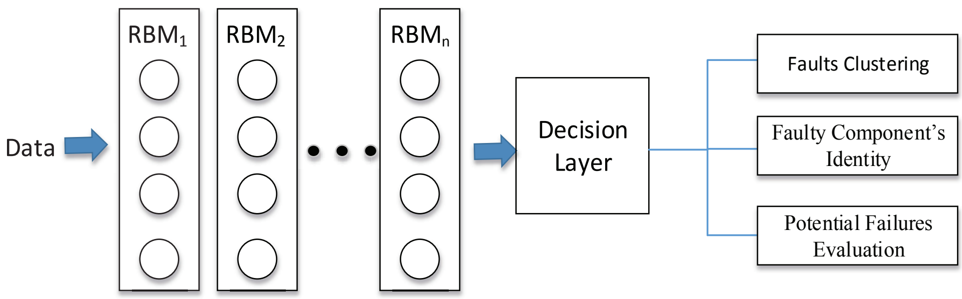

Figure 5.6 DBN for fault diagnosis and prognosis

Since DBN is based on RBM, which are particular energy-based models, the learning process would correspond to modifying that energy function so that its shape has desirable properties [Bengio, 2009]. In a DBN, each RBM is trained to encode in its weight matrix a probability distribution that predicts the activity of the visible layer through the hidden layer, which enables DNB the ability of self-learning. By stacking such models, and letting each layer predict the activity of the layer below, higher RBM learn increasingly abstract representations of sensory inputs [O'Connor et al., 2013]. A layer-by-layer nonlinear learning network with fine-tuning procedure enables DBN to capture intrinsic characteristics about potential failures from the massive data. Furthermore, energy-based models enable DBN to mine information hidden behind highly coupled inputs, which makes DBN a feasible method for fault diagnosis and prognosis when the target condition is beyond the historical data. Other applications of DBN for predictive maintenance are also listed in Chapter 2, Table 2.2. Next section will introduce and detail a novel application of DBN-based backlash error prediction in a machining center, when the target condition is beyond historical data.

5.5 Backlash error detection and prediction experiment

5.5.1 Experiment set up

In order to measure the backlash error in a machining center, at least two geometric measurements are required: displacement of the ball screw and the linear position of the spindle. In our experiment, the displacement of the ball screw is measured through a rotary encoder in the motor. It records the rotation angle and converts the signal into linear displacement. The position of the spindle is acquired by a calibrated linear scale. Figure 5.7 shows the setup of the measurement system, in which the linear scale records the direct position of the spindle x_1, and rotary encoder measures the displacement of the ball screw (the indirect position of the spindle x_2).

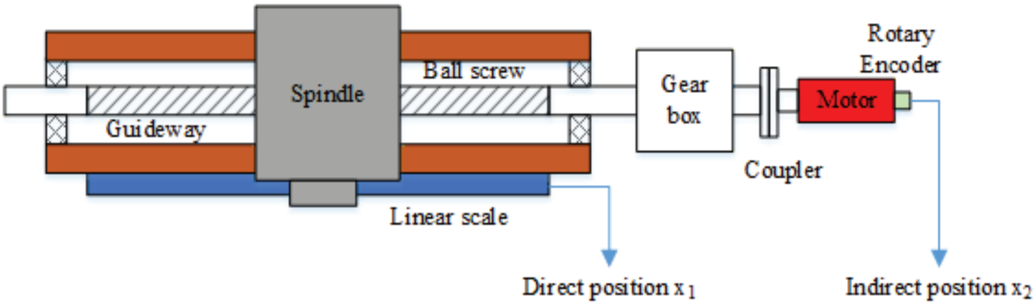

Figure 5.7 Setup of the measurement system

5.5.2 Backlash error interpretation

According to the definition, the geometric error caused by backlash error can be interpreted as the difference between the displacements of spindle and ball screw when the direction is reversed. Simply described as the following equations:

$$B[p] = \int_{t^b}^{t^*} f_1(t)\, dt - \int_{t^b}^{t^*} f_2(t)\, dt \tag{5.13}$$

$$p = x_1(t^b), \tag{5.14}$$

Where:

$f_1(t)$ is the velocity of the spindle with time, recorded by linear scale.

$f_2(t)$ is the velocity of the ball screw with time, recorded by rotary encoder.

$x_1(t)$ is the direct position of the spindle with time, recorded in linear scale.

p is the position on linear scale, where backlash occurs.

$B[p]$ is the geometric error caused by backlash at position p.

t^b is the time when backlash occurs.

t^* is the time when backlash disappears (the gap is taken occurs)

However, due to the uncertainty of backlash, it is almost impossible to capture the exact time when the gap is filled, which means t^* in the equation (5.13) is unavailable or with low precision. Therefore, during the measurement, we extended the measurement distance and let backlash occurred three times in one sample to eliminate the measurement error. As shown in Figure 5.8 three blocks of geometric error, caused by backlash, can be recognized during one sampling process. The time interval between each sampling units is two milliseconds.

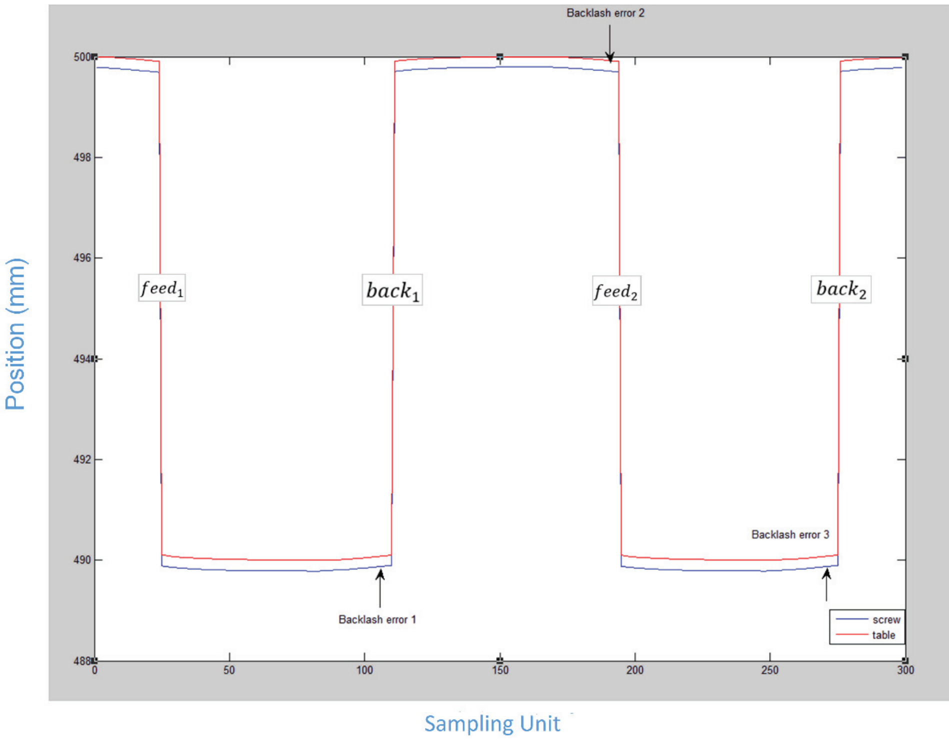

Figure 5.8 Three blocks of backlash in one sample

The position of spindle and ball screw can be decomposed into the amount of feed and return, the initial position, and backlash error, as shown in the following equation:

$$x_1(t') = x_1(t^0) + \int_{t^0}^{t'} feed(t)\,dt + \int_{t^0}^{t'} back(t)\,dt \;+ \sum backlash \tag{5.15}$$

$$x_2(t') = x_2(t^0) + \int_{t^0}^{t'} feed(t)\,dt + \int_{t^0}^{t'} back(t)\,dt \tag{5.16}$$

Where:

$x_2(t)$ is the position of ball screw with time, recorded in rotary encoder.

t^0 The initial time of the measurement.

t' The end time of the measurement.

$feed(t)$ The feed rate at t, which is equal to zero if not under feed movement.

$back(t)$ The return rate at t, which is equal to zero if not under return movement.

Since the backlash errors occur in very close positions, we can consider the difference of the value is approach to zero. Therefore, the backlash error can be calculated according to equation (5.17):

$$backlash = [x_1(t') - x_1(t^0)] - [x_2(t') - x_2(t^0)] \tag{5.17}$$

5.5.3 Data acquisition

To investigate the actual performance of proposed system, the experiment was carried out on a Pietro Carnaghi AC 16 TM vertical machining center with a collection of 25 weeks' data, from the 7th week to 31st since the last maintenance. The machining center is placed in a plant with some manufacturing tasks every day. During the data collection period, data was collected through a very rigorous testing procedure after daily work to ensure all the data collected is under a similar condition. The collected parameters are shown in Table 5.1.

During the experiment, we divided the moving distance into 24 points with the interval of 20 mm from 1090 to 1550 mm according to the calibrated linear scale. Then, all the backlash errors with current parameters can be calculated through backlash error interpretation method, which was introduced in Section 5.5.2. Figure 5.9 shows the obtained backlash error only with working weeks and axis position (It is more than a 3-dimension problem, but according to empirical knowledge, it is intuitive to visualize the backlash error with working weeks and axis position).

Table 5.1 Parameters collected from a vertical machining center

Parameters	Meaning
x_1	Direct position measured by linear scale

x_2	Displacement recorded by rotary encoder
w	Number of weeks since the last maintenance
T_1	Temperature of the coolant tank
T_2	Temperature of the machining center
T_3	Ambient temperature
TRQ	Machining torque
t	Sampling units during the testing procedure
P	Axis position of the spindle

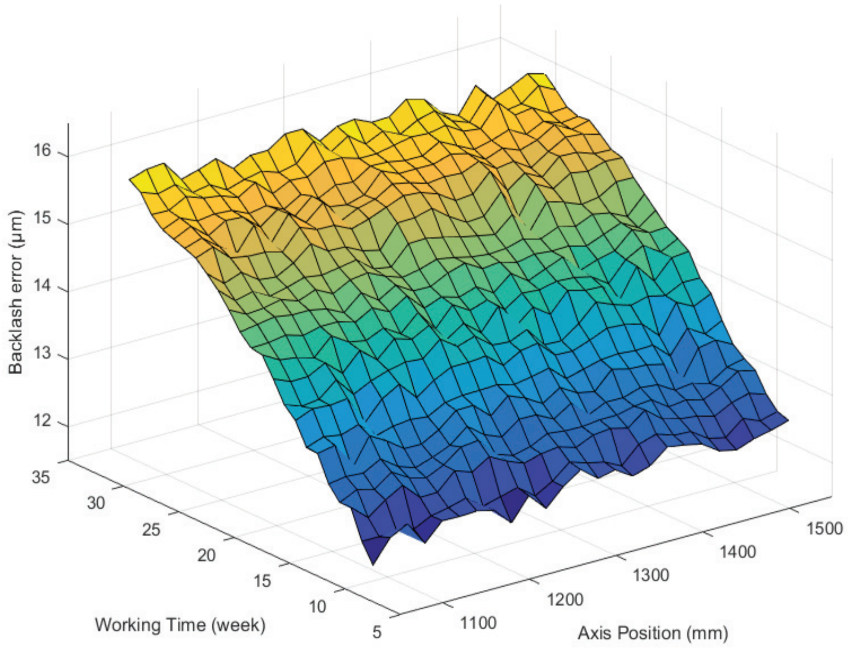

Figure 5.9 Backlash error with working time and axis position

5.5.4 Diagnosis of backlash error

As mentioned above, backlash errors in some positions may be missed in data acquisition layer. So Diagnosis Layer requires to detect and fill up the current backlash error in all positions through the collected parameters. The essence of the diagnosis model is to deal with a nonlinear regression problem, in which all the variables are collected or mapped under the same working condition. The regression function for the target backlash error can be simply shown as equation (5.18):

$$b_t \approx F(T_1, T_2, T_3, t, TRQ, p_r, b_r, p_t) \tag{5.18}$$

Where b_t is the backlash error at target position, p_r represents the backlash error at reference position, which is obtained and interpreted in data acquisition layer, p_t and b_r mean the target position and reference position, respectively.

In Diagnosis Layer, the diagnosis is first trained through the historical data, and responsible to map the backlash error in all positions through the data collected under current working condition. During the experiment, a selection of 1152 samples was applied to train the regression model. We employed three methods including BPNN, SVMR, and DBN as the regression model, and compared the results as shown in Table 5.2, where Maximum Error (ME), Mean Squared Error (MSE) and training time of each model are listed.

Table 5.2 Diagnosis results of BPNN, DBN, and SVMR

Model	BPNN	DBN	SVMR
Structure	50 nodes in the hidden layer	4 layers, 50 nodes in each	Gaussian Kernel Function
MSE (μm)	0.01148	0.01056	0.00964
ME (μm)	0.2758	0.2807	0.2812
Training time	Instantly	30 mins	Instantly

The result shows that all three methods have the capacity to deal with the diagnosis problem when all the relative parameters were obtained. In addition, both BPNN and SVMR can finish the training process instantly, and DBN is relatively time-consuming.

5.5.5 *Prognosis of backlash error*

As mentioned above, Prognosis Layer is responsible to predict future backlash error according to the historical data and current working condition of the machining center. Actually, it can also be considered as a regression problem, though all the parameters that can be used to represent the working condition are missing since we never know the exact values of these parameters in the future. Therefore, we applied the deep learning approach to predict the backlash error first, and then tested other methods to compare the effectiveness. Table 5.3 shows the input variables of the prognosis model.

Table 5.3 Inputs for prognosis

Inputs	Meaning
w	Number of weeks since the last maintenance
T_1	Temperature of coolant tank
T_2	Temperature of machine
T_3	Ambient temperature
TRQ	Machine torque
P	Axis position
$backlash_{w-1,P-1}$	Backlash error in $w-1$ weeks at $P-1$ position
$backlash_{w-2,P-2}$	Backlash error in $w-2$ weeks at $P-2$ position

During the experiment, it is supposed that prognosis model is established at week 29th and employed to predict the backlash error in the future, which means the backlash errors in week 30th and 31st are beyond the training data. According to the author's empirical knowledge, this problem is common and challenging since in many situations, one may not have the data under faults. Some machines may run several years without any failures. However, it may exist potential faults that would occur one day. When they happen, they may cause terrible disasters in both economy and personal safety. This is the reason why it is also crucial to evaluate potential failures or degradations beyond the historical data.

During the experiment, the applied DBN model is constructed through stacking four RBM. The structure for the utilized DBN is shown in Table 5.4 along with the parameters of each Restricted Boltzmann Machine.

Table 5.4 parameters for utilized DNB

Parameters	RBM1	RBM2	RBM3	RBM4
Type	Bernoulli	Bernoulli	Bernoulli	Bernoulli
Number of neurons	50	50	30	30
Learning rate	0.01	0.01	0.01	0.01
Number of epochs	500	500	300	300

During the training process, samples are randomly divided into two groups, 80 percent for training and 20 percent for testing. The training process has been run for 20 times, and there are no significant divergence among the results. Figure 5.10 shows one of these results, in which the best MSE from week 9th to week 29th is 0.012207 μm at the 14075th epochs.

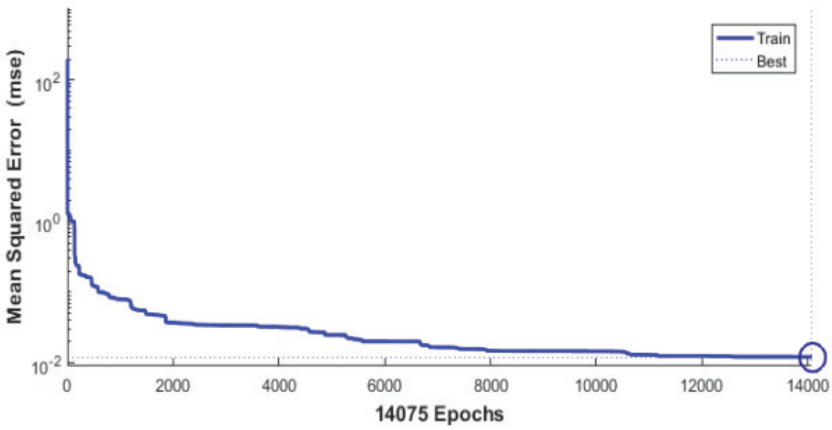

Figure 5.10 Training result of DBN

Figure 5.11 and Figure 5.12 compared the actual backlash error and predicted backlash error in week 30th and 31st respectively. The MSE of backlash error prediction in week 30th

and 31st is 0.0102 µm and 0.0142 µm with 0.1808 and 0.2275 as ME, respectively. In this case, the Maximum Permissible Error (MPE) for the backlash error is set as 16µm. In week 29th, we can predict that the backlash error in week 31st may exceed the MPE considering about the mean prediction error, which means the fault could be forecasted two weeks in advance. Then, subsequent maintenance must be arranged at or before week 30th to prevent the error from growing up and finally exceeding the MPE.

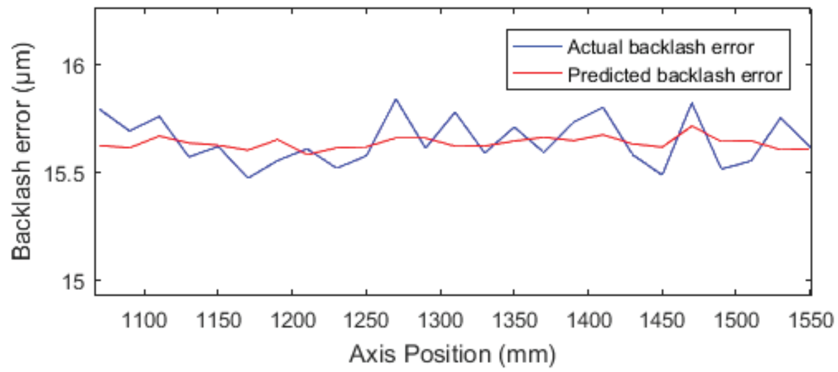

Figure 5.11 Predicted backlash error in week 30th

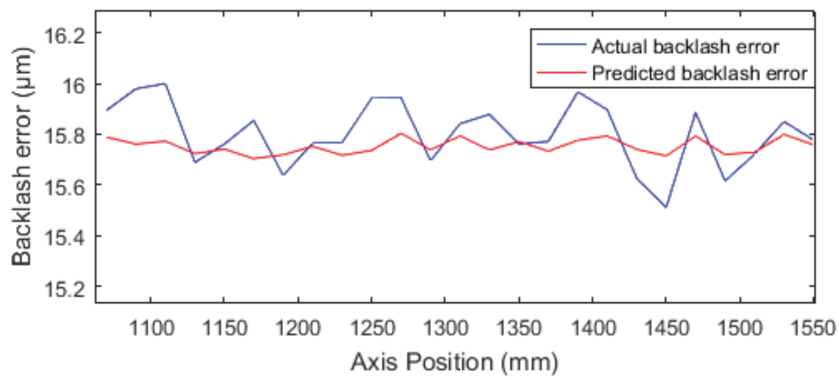

Figure 5.12 Predicted backlash error in week 31st

5.5.6 Discussion

During the experiment, to confirm the effectiveness of deep learning for backlash error prediction, two other widely used regression methods, SVMR and BPNN, are employed for comparison. The BPNN applied in the experiment consisted of one hidden layer with 50 neurons. The Levenberg-Marquard optimization rule is applied as training algorithm for the BPNN. For the SVMR, Gaussian Function is applied as the kernel function. As mentioned above, all the training samples are randomly selected. In order to obtain more accurate and comprehensive evaluation results, we did this random selection for 20 times. The prediction results of each algorithm with their average performance over the 20 runs are summarized in Table 5.5.

Table 5.5 Prediction results of BPNN, DBN, and SVMR

Model	BPNN	DBN	SVMR
Structure	50 nodes in the hidden layer	4 layers, 50 * 50 * 30 * 30	Gaussian Kernel Function
training MSE (μm)	0.020894	0.012207	0.0102
MSE in week 30th	0.2566	0.0102	1.2335
MSE in week 31st	1.6450	0.0142	2.8029
ME in week 30st	0.9904	0.1808	1.3669
ME in week 31st	1.6338	0.2275	1.8969
Training time	Instantly	40 mins	instantly

Through observation, DBN performs much better than other two methods with high accuracy in both MSE and ME for backlash error prediction, which indicates that DBN can effectively deal with the backlash error prediction issue, when the important parameters are missing and the objective condition is beyond the training data. To be more specific, energy-based models enable DBN to mine information hidden behind highly coupled inputs, which makes DBN a feasible method to predict backlash error in the future through current condition. However, compared to the other two methods, DBN is a kind of time-consuming approach. And the training time may increase dramatically with the growth of training samples due to the complex structure.

As to BPNN and SVMR, both of them have good performances in fault diagnosis if all the necessary parameters can be obtained. In addition, compared with DBN, they showed the advantage in learning speed during the experiment. However, it is also obvious that neither of them has the ability to predict the backlash error in the future when the objective condition is beyond the training data. The case study also demonstrated the necessity and feasibility of applying DBN as the prognosis model for backlash error prediction in the proposed HDPS.

5.6 HDPS-BPSO maintenance implementation strategy

In order to capture the trade-off between several factors such as maintenance cost, machining accuracy, and defective percentage, a novel HDPS-BPSO maintenance implementation strategy driven by HDPS and binary particle swarm optimization (BPSO) is proposed in this section. After discovering fault information of the equipment, the last step is to implement predictive maintenance according to the prediction of potential failures or degradation, which is usually a NP-hardness (non-deterministic polynomial-time hardness) problem. Here, the implementation strategy can be regarded as a maintenance scheduling optimization problem. Inspired by particle swarm optimization's (PSO) advantages [Rini et al., 2011], a novel HDPS-BPSO maintenance implementation strategy is proposed to find the optimum solution for predictive maintenance implementation. Since PSO is easier to implement with a few parameters to tune and is computationally inexpensive [Yue-Jiao et al., 2012], it may be a perfect solution in this case.

5.6.1 Basis of PSO

Particle swarm optimization is a computational method to solve the optimization problems, by iteratively trying to improve candidate solutions with communication within the swarm and randomly search, which is inspired from movement of organisms in a bird flock [R. Eberhart and Kennedy, 1995], as shown in Figure 5.13. The current position $\overrightarrow{x_i}$ can be considered as a set of coordinates describing a point in space. If the current position is better than any that has been found so far, then the coordinates are stored in the vector $\overrightarrow{p_i}$. The value of the best function result so far is stored in a variable that can be called $\overrightarrow{p_g}$. The objective, of course, is to keep finding better positions and updating $\overrightarrow{p_i}$ and

$\overrightarrow{p_g}$. New points are chosen by adding $\overrightarrow{v_i}$ coordinates to $\overrightarrow{x_i}$, and the algorithm operates by adjusting $\overrightarrow{v_i}$, which can effectively be seen as a step size.

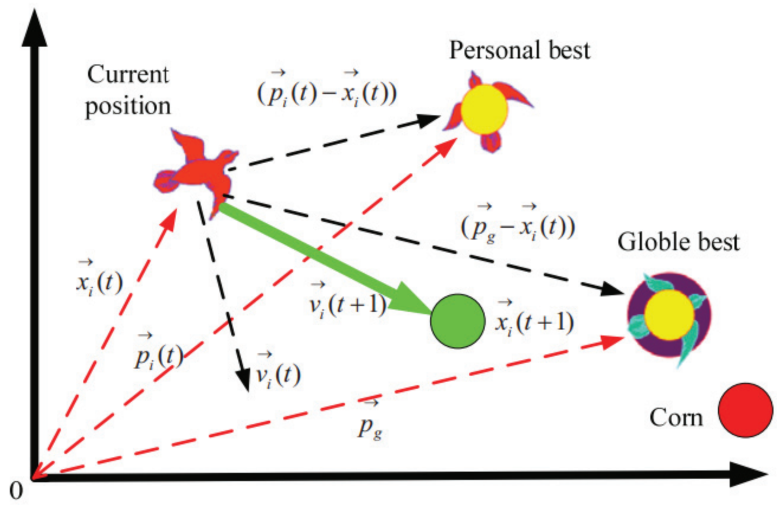

Figure 5.13 Birds flocking of PSO

In optimization process, a population of candidate solutions are produced in the form of particles. These particles move around in the solution space of the problem according to some simple mathematical formulae over the particle's positon and velocity. The movement of each particle is influenced by the best known personal position and also the best known global position in the searching space, which is updated as the best solution found so far by the swarm. This update makes the swarm move toward the best solutions [Q. Yu, 2015]. The flowchart of PSO can be seen in Figure 5.14.

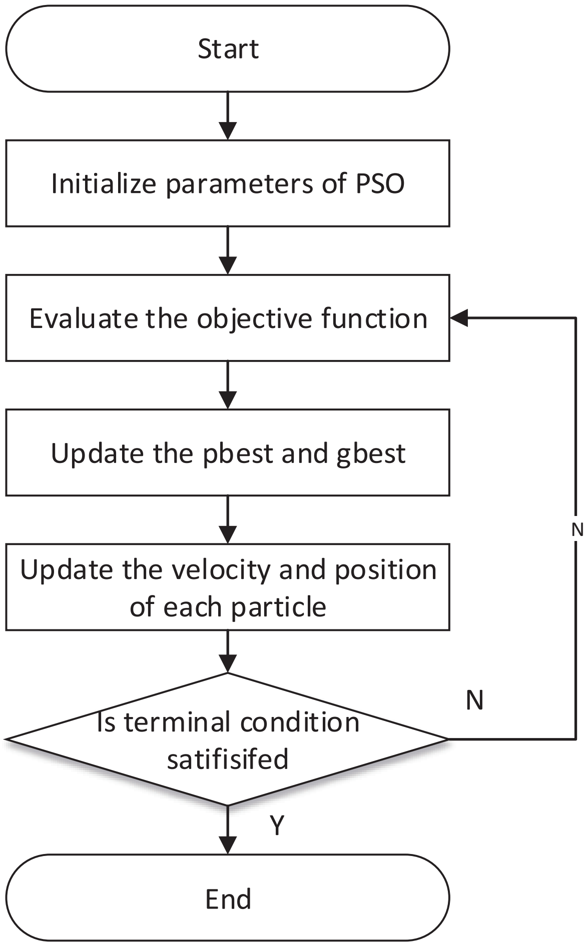

Figure 5.14 Flowchart of PSO algorithm

In PSO, every particle remembers its own previous best value as well as the neighborhood best. PSO is also more efficient in maintaining the diversity of the swarm, since all the particles use some information related to the most successful particle in order to improve themselves. In addition, PSO is easier to implement and there are only a few parameters to adjust. The general steps of implementing PSO were shown as follows:

1. Initialize parameters such as maximum number of iterations, population size and initial

particle velocities $\vec{v}_i$ and positions $\vec{x}_i$.

2. Loop

3. Evaluate the target optimization fitness according to each particle's position $\vec{x}_i$

4. Update the best solution of each particle $\vec{p}_i$ so far.

5. Update the best solution of all particles $\vec{p}_g$ until now.

6. Change the velocity of each particle at t^{th} iteration to $(t+1)^{th}$ iteration according to:

$$\vec{v}_i(t+1) = \omega \cdot \vec{v}_i(t) + c_1 r_1 \left(\vec{p}_i - \vec{x}_i(t)\right) + c_2 r_2 \left(\vec{p}_g - \vec{x}_i(t)\right) \tag{5.19}$$

Where w is the inertia weighting, $c_1$and c_2 are acceleration coefficients and $r_1$and r_2 are random numbers distribution on [0, 1].

7. Update the position of each particle according to the following equation:

$$\vec{x}_i(t+1) = \vec{x}_i(t) + \vec{v}_i(t+1) \tag{5.20}$$

8. If a criterion is met, exit loop. The criterion is usually set to be the maximum iterations, the number of iterations in which the objective has not been improved, or the fitness is sufficiently good.

The role of inertia weight w in Equation 5.19 is considered critical for the convergence behavior of PSO. The inertia weight is employed to control the impact of the previous history of velocities on the current one. Accordingly, the parameter w regulates the trade-off between the global (wide-ranging) and local(nearby) exploration abilities of the swarm. A large inertia weight facilitates global exploration, i.e. searching new areas, while a small one tends to facilitate local exploration, i.e. fine-tuning the current search area. A suitable value for the inertia weight w usually provides balance between global and local exploration abilities and consequently results in reduction of the number of interactions required to locate the optimum solution. Initially, the inertia weight is set as a constant.

However, some experiment results indicate that it is better to initially set the inertia to a large value, in order to promote global exploration of the search space, and gradually decrease it to get more refined solutions[R. C. Eberhart and Shi, 2000]. Thus, an initial value is set to maximum one w_{max} (for example around 1.2) and gradually reducing towards the minimum one w_{min} (for example around 0.6) can be considered as a good choice. A better method is to use some adaptive approaches, in which the parameters can be adaptively fine-tuned according to the problems under consideration [Hongbo Liu et al., 2007; Shi and Eberhart, 2001].

The parameters $c_1$and c_2 in Equation 5.19 are not critical for the convergence of PSO. However, proper fine-tuning may result in faster convergence and alleviation of local minima. As default values, usually, $c_1 = c_2$ =2 are used, but some experiment results indicate that $c_1 = c_2$ =1.49 might provide even better results. According to Equation 5.19, it is better for local exploration when $c_1 > c_2$, while global exploration would do better when $c_1 < c_2$. Some research also reports that it might be even better to choose a larger cognitive parameter, c_1, than a social parameter, c_2, but with $c_1 + c_2 < 4$ [Clerc and Kennedy, 2002]. Therefore, the parameter c_1 can be changed from c_{1min} to c_{1max} and the parameter c_2 can be changed from c_{2max} to c_{2min} regularly in order to make the algorithm promote global exploration in the beginning and get more refined solutions (local exploitation) in the end [Zhang, 2014].

5.6.2 BPSO

PSO was originally developed for continuous valued spaces [Khanesar et al., 2007], however, many practical problems are defined for discrete valued spaces where the domain of the variables is finite. In 1997, Kennedy and Eberhart proposed a discrete binary version of PSO for discrete optimization problems [Kennedy and Eberhart, 1997]. In their model, each particle represents its position in binary values which are 0 or 1. Each particle's value can then be changed from one to zero or vice versa.

In BPSO, the particle's personal best and global best are also updated as in continuous version. The main difference lies on the moving velocity, which is defined in terms of changes of probabilities that a bit will be in one state or the other. Therefore, velocity must be restricted within the range [0,1] through defining a logistic transformation S, usually a sigmoid function as Equation (5.21).

$$S\left(v_{ij}(t)\right) = \frac{1}{1+e^{-v_{ij}(t)}} \quad (5.21)$$

Where $v_{ij}(t)$ means the j^{th} component of vector $\vec{v_i}(t)$. Then the new position of the particle could be updated according to Equation (5.22).

$$if\ rand_{ij} < S\left(v_{ij}(t)\right)\ then\ x_{ij}(t+1) = 1 \quad (5.22)$$

$$otherwise\ x_{ij}(t+1) = 0$$

Where $rand_{ij}$ is a random number selected from a uniform distribution in [0, 1], $x_{ij}(t+1)$ represents the j^{th} component of vector $\vec{x_i}(t+1)$.

However, as reported in [Nezamabadi-pour et al., 2008], increasing the value in the positive direction in the BPSO will cause larger probability (probability of 1) for the particle position while raise in the negative direction results in probability of zero. When the optimization process has nearly reached to the optimum solution, the probability of changing the position of the particle must be near to zero, while at this point using sigmoid function, the position will change by taking the value of 1 or 0 with the probability of 0.5, which would cause the algorithm not to converge well. To avoid this situation, hyperbolic tangent (Tanh) function, as shown in Equation (5.23), is leveraged as the transformation function

$$S\left(v_{ij}(t)\right) = \left|tanh\left(\alpha v_{ij}(t)\right)\right| = \frac{e^{\alpha v_{ij}(t)} - e^{-\alpha v_{ij}(t)}}{e^{\alpha v_{ij}(t)} + e^{-\alpha v_{ij}(t)}} \quad (5.23)$$

Where α is the weight vector of the transportation.

5.6.3 HDPS-BPSO based maintenance scheduling

As introduced above, backlash error that will occur in the equipment at all positions and directions could be predicted through proposed HDPS. Table 5.6 shows part of backlash errors (μm) in the machining center at x direction predicted through HDPS. During the scheduling, our target is to minimize the total cost raised by backlash error, including the maintenance cost, machining accuracy, and defective percentage in the latest 25 weeks. The specific data about work load of the equipment is shown in Table 5.7 (%).

Table 5.6 Part of data predicted through HDPS (μm)

	Position A	Position B	Position C	Position D
Week10	12,35	12,35	12,37	12,35
Week11	12,44	12,47	12,51	12,41
Week12	12,64	12,64	12,77	12,74
Week13	12,84	12,80	12,85	12,97
Week14	12,98	12,93	12,98	13,02
Week15	13,22	13,12	13,11	13,19
Week16	13,30	13,36	13,38	13,36
Week17	13,46	13,62	13,71	13,56
Week18	13,77	13,83	13,87	13,92
Week19	13,99	14,13	14,01	14,14
Week20	14,32	14,21	14,20	14,25
Week21	14.42	14.43	14.39	14.46
Week22	14.59	14.56	14.50	14.61

Table 5.7 Work load of the equipment (%)

Week	Load	Week	Load	Week	Load
1	91.11	10	93.33	19	86.67
2	97,78	11	82,22	20	84,44
3	95.56	12	77,78	21	80,0
4	88.89	13	84,44	22	88,89
5	82.22	14	88,89	23	80
6	80	15	82,22	24	84,44
7	82,22	16	88,89	25	77,78
8	95.56	17	84,44		
9	84,44	18	91.11		

The main cost function in this case study includes degradation cost C_G, maintenance cost C_M , and inspection cost C_I. Here, the assumptions and definitions in the mathematical model are given:

Assumption 1: As we discussed in Chapter 3, in practical industrial applications, the relationship between production and maintenance is usually considered as a conflict in management decision. Here, we assume the maintenance scheduling compromises the production scheduling, which means the work load of the equipment will not change with maintenance decisions.

Assumption 2: The degradation in specific direction and position completely follows the mapping provided by HDPS.

Assumption 3: Once a maintenance has been performed, the degradations in all directions and positions are supposed to return back to the initial values (Week 1). Subsequent degradations keep following HDPS according to the distance from the last maintenance performance.

Assumption 4: If a maintenance has been scheduled, it is supposed to be performed at the beginning of that week.

Assumption 5: Holidays have been excluded from the mathematical model.

C_G: Degradation cost.

C_M: Maintenance cost.

C_I: Inspection cost.

W: Number of weeks to be scheduled.

A: Number of axes inspected.

P: Number of axial positions inspected.

Pr: Production profit in unit time.

D_P: Maximum permissible degradation.

D_N: Criterion of normal product.

M: Cost of maintenance performance.

$Load_i$: Working load in week i.

H: Maximum working hours per week.

h: Time of single maintenance performance.

D_{ijk}: Degradation in week i along j axis at position k predicated from HDPS.

D'_{ijk}: Degradation in week i along j axis at position k after maintenance scheduling.

α: Weighting factor for degradation cost.

β: Weighting factor for maintenance cost.

d_i: Distance from the last maintenance in week i.

x_i: Decision variable.

Decision variable x_i during optimization is defined as:

$$\begin{aligned} &if\ maintenance\ performanced\ in\ week\ i \quad then\ \ x_i = 1 \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad otherwise\ x_i = 0 \end{aligned} \tag{5.24}$$

The degradation cost here is caused by the geometrical error from backlash directly. It could be estimated as following:

$$C_G = \sum_{i\in W}\sum_{j\in D}\sum_{k\in P} Load_i * H * \varphi(D'_{ijk}) \tag{5.25}$$

$$D'_{ijk} = D_{d_i jk}$$

$$\begin{aligned} &if \quad x_i = 1 \qquad then\ d_i = 1 \\ &\qquad\qquad\qquad\quad otherwise\ d_i = d_{i-1} + 1 \end{aligned}$$

Where $\varphi(\ )$ denotes the production cost caused by degradation. It can be calculated according to Equation (5.26):

$$\varphi(D'_{ijk}) = \begin{cases} 0 & if\ \ D'_{ijk} \le D_N \\ \frac{D'_{ijk} - D_N}{D_P} * Pr & if\ \ D_N < D'_{ijk} \le D_P \\ Pr & if\ \ D'_{ijk} > D_P \end{cases} \tag{5.26}$$

Here, we consider when the degradation is between the normal and maximum permissible degradation, the manufacturing profit decrease with a linear manner with degradation. The maintenance cost here is evaluated according to the number of maintenance performance.

$$C_M = M * \sum_{i\in W} x_i \tag{5.27}$$

Then, the total cost C_{tot}can be obtained as:

$$C_{tot} = \alpha * C_G + \beta * C_M + C_I \tag{5.28}$$

With constraint $\forall x_i \in W\colon\ x_i * h + Load_i * H \le H$

Because the equipment is inspected in a continuous manner in this model, the value of C_I is fixed. Since some issues such as incidental damage or cost caused by maintenance, and the loss in reputation of producing imperfect products. α and β can be leveraged to weight the effect of degradation and maintenance here, respectively.

The parameters of HDPS-BPSO are set according to the case study as: number of population size is 100, maximum iteration is 500, weighting coefficients α and β are both set as 1, W is 25 weeks, A is 2 axes, P is 25 positions, Pr is 2,000 NOK /hrs, M is 15,000 NOK, D_P is 16 µm, D_N is 12.5 µm, H is 45 hours, h is 2 hours. During the test, we leveraged hyperbolic tangent function as logistic transformation for optimization. The numerical result of HDPS-BPSO is as following. Figure 5.15 illustrates the mean fitness during the optimization. The convergence starts around 200^{th} iteration. Table 5.8 shows the optimum maintenance implementation scheduling according to proposed HDPS-BPSO. It means the best predictive maintenance solution in this case is to perform maintenance in week 9 and week 18, in which the total cost including the loss from degradation and maintenance cost is 33,303 NOK. According to the previous preventive maintenance strategy, the maintenance is supposed to be performed every 6 weeks. The cost is also calculated based on the preventive maintenance strategy. When maintenance executed in week 7, 13 and 19. The total cost is 47,881 NOK. Therefore, through predictive maintenance, the maintenance cost of single machine center can be reduced by 14,578 NOK in this case.

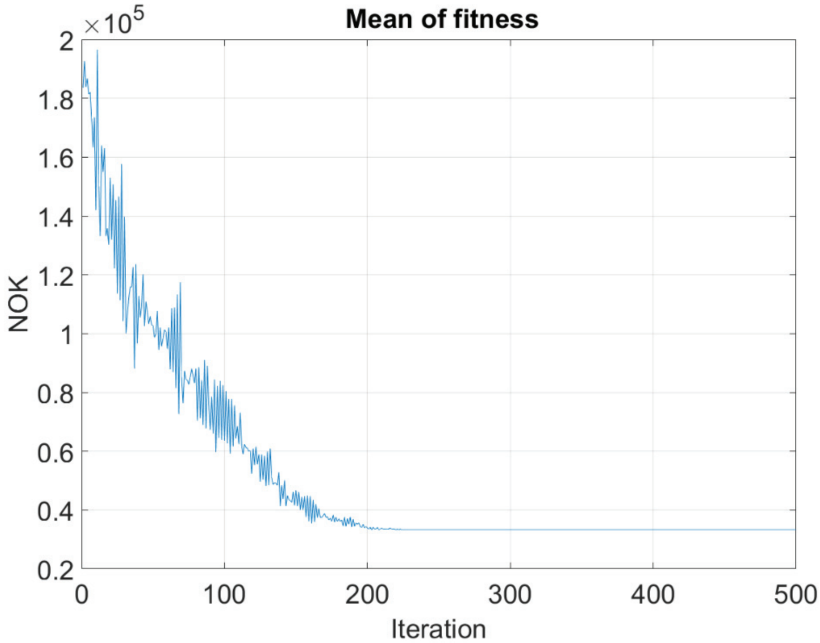

Figure 5.15 HDPS- BPSO mean fitness with iterations

Table 5.8 Best maintenance scheduling from HDPS-BPSO

Strategy	Criterion	Decision	Cost
Preventive maintenance	Time-based	Week 7,13,19	47,881 NOK
Predictive maintenance	Cost minimization	Week 9, 18	33,303 NOK

5.7 Summary

In this chapter, a case study of applying deep learning to predict backlash error for maintenance implementation scheduling in a machining center is demonstrated. HDPS is proposed for backlash error detection and prediction based on DBN to deal with the situation when target condition is beyond the historical data. The case study demonstrated the performance of HDPS for the backlash error prediction in a vertical machining center. During diagnosis stage, the missing prior data including the historical data and current backlash error will be interpreted, while the prognosis stage is responsible for the prediction of future backlash error based on the prior data provided by the former stage through a deep neural network. To provide a comprehensive comparison for the effectiveness of HDPS, two other intelligent algorithms, BPNN and SVMR, are also applied to replace the DBN as the prognosis model. The result of the comparison proves the superiority to apply deep learning method for backlash error prediction, when important parameters are missing and the objective condition is beyond the training data. Moreover, a novel maintenance implementation strategy HDPS-BPSO is also proposed to illustrate the implementation of predictive maintenance in practical application. The numerical result also shows the benefit of implementing the strategy of predictive maintenance compared with that of preventive maintenance

References

Bengio, Y. (2009). Learning deep architectures for AI. *Foundations and trends® in*

Machine Learning, 2(1), 1-127.

Carreira-Perpinan, M. A., and Hinton, G. (2005). *On Contrastive Divergence Learning.* Paper presented at the AISTATS.

Chen, C., Liu, Z., Zhang, Y., Chen, C. P., and Xie, S. (2016). Actuator Backlash Compensation and Accurate Parameter Estimation for Active Vibration Isolation System. *IEEE Transactions on Industrial Electronics, 63*(3), 1643-1654.

Cheng, Q., Zhao, H., Zhang, G., Gu, P., and Cai, L. (2014). An analytical approach for crucial geometric errors identification of multi-axis machine tool based on global sensitivity analysis. *The International Journal of Advanced Manufacturing Technology, 75*(1-4), 107-121.

Clerc, M., and Kennedy, J. (2002). The particle swarm-explosion, stability, and convergence in a multidimensional complex space. *IEEE transactions on Evolutionary Computation, 6*(1), 58-73.

Eberhart, R., and Kennedy, J. (1995). *A new optimizer using particle swarm theory.* Proceedings of the Sixth International Symposium on the Micro Machine and Human Science, 1995. MHS'95.

Eberhart, R. C., and Shi, Y. (2000). *Comparing inertia weights and constriction factors in particle swarm optimization.* Paper presented at the Evolutionary Computation, 2000.

Erhan, D., Bengio, Y., Courville, A., Manzagol, P.-A., Vincent, P., and Bengio, S. (2010). Why does unsupervised pre-training help deep learning? *Journal of Machine Learning Research, 11*(Feb), 625-660.

Fines, J. M., and Agah, A. (2008). Machine tool positioning error compensation using artificial neural networks. *Engineering Applications of Artificial Intelligence, 21*(7), 1013-1026.

Gan, M., and Wang, C. (2016). Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings. *Mechanical Systems and Signal Processing, 72*, 92-104.

Hinton, G. (2010). A practical guide to training restricted Boltzmann machines. *Momentum,* *9*(1), 926.

Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A.-r., Jaitly, N., . . . Sainath, T. N. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine, 29*(6), 82-97.

Hinton, G. E. (2002). Training products of experts by minimizing contrastive divergence. *Neural computation, 14*(8), 1771-1800.

Hinton, G. E., Osindero, S., and Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. *Neural computation, 18*(7), 1527-1554.

Huang, N., Bi, Q., and Wang, Y. (2015). Identification of two different geometric error definitions for the rotary axis of the 5-axis machine tools. *International Journal of Machine Tools and Manufacture, 91*, 109-114.

Jiang, X., and Cripps, R. J. (2015). A method of testing position independent geometric errors in rotary axes of a five-axis machine tool using a double ball bar. *International Journal of Machine Tools and Manufacture, 89*, 151-158.

Kao, J., Yeh, Z.-M., Tarng, Y., and Lin, Y. (1996). A study of backlash on the motion accuracy of CNC lathes. *International Journal of Machine Tools and Manufacture, 36*(5), 539-550.

Kennedy, J., and Eberhart, R. C. (1997). *A discrete binary version of the particle swarm algorithm.* Paper presented at the Systems, Man, and Cybernetics, 1997. 1997 IEEE International Conference on Computational Cybernetics and Simulation.

Keyvanrad, M. A., and Homayounpour, M. M. (2014). A brief survey on deep belief networks and introducing a new object oriented MATLAB toolbox (DeeBNet V2. 1). *arXiv preprint arXiv:1408.3264*.

Khanesar, M. A., Teshnehlab, M., and Shoorehdeli, M. A. (2007). *A novel binary particle swarm optimization.* Paper presented at Mediterranean Conference on the Control & Automation 2007. MED'07.

Kusiak, A. (2017). Smart manufacturing must embrace big data. *Nature, 544*(7648), 23.

Lee, K.-I., and Yang, S.-H. (2013). Measurement and verification of position-independent geometric errors of a five-axis machine tool using a double ball-bar. *International Journal of Machine Tools and Manufacture, 70*, 45-52.

Li, J., Tao, F., Cheng, Y., and Zhao, L. (2015). Big data in product lifecycle management. *The International Journal of Advanced Manufacturing Technology, 81*(1-4), 667-684.

Li, Z., Wang, Y., and Wang, K. (2017). A data-driven method based on deep belief networks for backlash error prediction in machining centers. *Journal of Intelligent Manufacturing*. doi:10.1007/s10845-017-1380-9

Liu, H., Abraham, A., and Zhang, W. (2007). A fuzzy adaptive turbulent particle swarm optimisation. *International Journal of Innovative Computing and Applications, 1*(1), 39-47.

Liu, H., Xue, X., and Tan, G. (2010). Backlash error measurement and compensation on the vertical machining center. *Engineering, 2*(06), 403.

Mosallam, A., Medjaher, K., and Zerhouni, N. (2016). Data-driven prognostic method based on Bayesian approaches for direct remaining useful life prediction. *Journal of Intelligent Manufacturing, 27*(5), 1037-1048.

Mourtzis, D., Vlachou, E., Milas, N., and Xanthopoulos, N. (2016). A cloud-based approach for maintenance of machine tools and equipment based on shop-floor monitoring. *Procedia CIRP, 41*, 655-660.

Nezamabadi-pour, H., Rostami-Shahrbabaki, M., and Maghfoori-Farsangi, M. (2008). Binary particle swarm optimization: challenges and new solutions. *CSI J Comput Sci Eng, 6*(1), 21-32.

O'Connor, P., Neil, D., Liu, S.-C., Delbruck, T., and Pfeiffer, M. (2013). Real-time classification and sensor fusion with a spiking deep belief network. *Frontiers in Neuroscience, 7*, 178. doi:10.3389/fnins.2013.00178

Prasanga, D. K., Tanida, K., Mizoguchi, T., and Ohnishi, K. (2013). *Evaluation of a backlash compensation method using two parallel thrust wires.* Paper presented at 2013 IEEE International Symposium on the Industrial Electronics (ISIE).

Rini, D. P., Shamsuddin, S. M., and Yuhaniz, S. S. (2011). Particle swarm optimization: technique, system and challenges. *International Journal of Computer Applications, 14*(1), 19-26.

Schwenke, H., Knapp, W., Haitjema, H., Weckenmann, A., Schmitt, R., and Delbressine, F. (2008). Geometric error measurement and compensation of machines—an update. *CIRP Annals-Manufacturing Technology, 57*(2), 660-675.

Shi, Y., and Eberhart, R. C. (2001). *Fuzzy adaptive particle swarm optimization.* Proceedings of the 2001 Congress on the Evolutionary Computation.

Siguenza-Guzman, L., Saquicela, V., Avila-Ordóñez, E., Vandewalle, J., and Cattrysse, D. (2015). Literature review of data mining applications in academic libraries. *The Journal of Academic Librarianship, 41*(4), 499-510.

Slamani, M., Nubiola, A., and Bonev, I. A. (2012). Modeling and assessment of the backlash error of an industrial robot. *Robotica, 30*(07), 1167-1175.

Tao, F., Cheng, J., Qi, Q., Zhang, M., Zhang, H., and Sui, F. (2017). Digital twin-driven product design, manufacturing and service with big data. *The International Journal of Advanced Manufacturing Technology*, 1-14.

Yu, D., and Deng, L. (2011). Deep learning and its applications to signal and information processing [exploratory dsp]. *IEEE Signal Processing Magazine, 28*(1), 145-154.

Yu, Q. (2015). New Approaches for Automated Intelligent Quality Inspection System Integration of 3D Vision Inspection, Computational Intelligence, Data Mining and RFID Technology.

Yue-Jiao, G., Meie, S., Jun, Z., Kaynak, O., Wei-Neng, C., and Zhi-Hui, Z. (2012). Optimizing RFID Network Planning by Using a Particle Swarm Optimization Algorithm With Redundant Reader Elimination. *IEEE Trans. Ind. Appl., 8*(4), 900-912. doi:10.1109/TII.2012.2205390

Zhang, Z. (2014). *Data Mining Approaches for Intelligent Condition-based Maintenance: A Framework of Intelligent Fault Diagnosis and Prognosis System (IFDPS).* NTNU-trykk.

Zhong, G., Wang, C., Yang, S., Zheng, E., and Ge, Y. (2015). Position geometric error modeling, identification and compensation for large 5-axis machining center prototype. *International Journal of Machine Tools and Manufacture, 89*, 142-150.

Zhu, S., Ding, G., Qin, S., Lei, J., Zhuang, L., and Yan, K. (2012). Integrated geometric error modeling, identification and compensation of CNC machine tools. *International Journal of Machine Tools and Manufacture, 52*(1), 24-29.

Chapter 6
Implementation of predictive maintenance in rotary machinery

Mechanical degradation may cause equipment to break down with serious safety, environment, and economic impact. Since rotary machinery usually operates under a tough working environment, which makes it vulnerable to types of faults and increases the complexity of fault diagnosis. Degradation assessment of components in mechanical equipment is usually unsatisfied or restricted by its accuracy in most cases. Simultaneously, the requirement of manufacturing systems with reliable self-assessment has been increasingly raised with the trend of smart industry. The aim of this chapter is to fill this gap by providing a deep learning driven method for fault classification and degradation assessment. An experiment for fault classification and degradation assessment in rotary machinery through wavelet packet decomposition (WPD) and data-driven models will be demonstrated in this chapter. During the experiment, WPD is first applied to represent the coefficient and energy based features from vibration signals. Then several machine-learning methods, including DNN, DBN, BPNN, SVM, and K-Nearest neighbour classification are leveraged for fault classification and degradation assessment. The comparison of numerical results shows the superiority of DNN for degradation assessment in rotary machinery. In addition, a novel SAE-LSTM approach will also be presented for anomaly detection through multiple features sequence when the history data is unlabelled, which is also a common dilemma in practical applications.

6.1 Introduction

As the key equipment in many production fields, rotary machinery covers a very broad range of industrial equipment and plays a momentous role in manufacturing application. It is one of the most common classes of mechanical equipment and generally may operate under a tough working environment, which make it vulnerable to types of faults. These faults may cause equipment to break down or degrade certain machinery performance like geriatric location, manufacturing quality and operation safety [Lei et al., 2013]. Considering the complexity of the current industrial applications, degradation assessment in machinery is a challenging issue nowadays [El Kadiri et al., 2016; Precup et al., 2015]. Studies have shown that the human operator is responsible for 70–90% of the accidents in industrial environments [P. Wang and Guo, 2013]. For this reason, computer-based degradation assessment systems with high complexity is imperative to improve the accuracy of fault identification, and prevent unanticipated accidents.

Moreover, rotary machinery is often critical to the ability of a production process to perform as and when required. Failure in such rotary machinery can have serious safety, environment, and economic impact. The aim of maintenance in rotary machinery usually lies on preventing the equipment from failures and reduce maintenance costs by decreasing the number of unnecessary maintenance. When degradation in rotary machinery has reached a point where it can be heard in form of noise, felt in form of heat, or seen in form of smoke, failure is about to materialise. Therefore, choosing the suitable inspection methodology to evaluate the degradation in rotary machinery with long warning time is of utmost importance. Simultaneously, vibration monitoring is widely leveraged as a main monitoring method for early detection of degradation in rotating machinery due to the good performance in representing fault information. However, in most cases, the subsystems in rotary machineries like bearings and gear transmission systems are not easily accessible, or hard to inspect visually the failures directly due to restrictions of time consuming disassembly, huge machine size or environmental limitations [Y. Yang et al., 2015]. Therefore, how to achieve early fault detection, classification and degradation assessment failures in rotary machinery is always a hot issue in the field of mechanical maintenance.

The research target usually focuses on the root cause of increased vibration levels. When the root cause, usually degradations on certain parts, is known, the right operation and maintenance action could be planned. Many intelligent approaches for diagnosis or prognosis in rotary machine have been proposed and researched in the recent years [Lin and Chen, 2014; Lu et al., 2017; Z.-Y. Wang et al., 2017]. Lin et al. [2014] applied crossover characteristics to extract failure features from nonlinear data to detect faults for rotary machine. Wang et al. [2017] proposed a method to selective ensemble neural networks for faults classification in rotary machine. Lu et al. [2017] introduced a stacked denoising autoencoder to estimate the health condition of rotary machinery components. All these proposed methods have contributed greatly and achieve certain targets in relevant experiment. However, in most cases, the accuracy of degradation assessment for certain components or performance in mechanical equipment is still unsatisfied due to the increasing complexity of current industrial applications. Therefore, this chapter proposes a method driven by WPD and DNN for fault classification and degradation assessment in rotary machinery.

6.2 Vibration condition monitoring

The art of anticipating failure in rotary machinery by means of monitoring vibration is widely used in industry as almost 80% of common rotating equipment problems relates to misalignment and imbalance are detectable by vibration monitoring. The measured vibration levels will change when a rotary machine has a defect. Vibrations caused by the defects occur at specific vibration frequencies, characteristic of the components, their operation, assembly and wear. The measured vibration levels could indicate the severity of the defects [Scheffer and Girdhar, 2004]. This may probably be the reason why vibration condition monitoring is the most popular method for faults identification and classification in rotary machinery.

Evaluation criteria for machine vibration is dependent upon a wide range of factors and the criteria adopted vary significantly for different types of machine. According to ISO standard 20816, there are three primary vibration quantities: displacement, velocity and acceleration ["ISO 20816-1," 2016]. However, in most situation, it is hard to give absolute vibration tolerances for any given machine. There is thus an obvious risk of judging measured vibration levels, dangerous when they are not, or the opposite - not dangerous when they are dangerous. Human experience and interpretation of the measured values still plays an important role in vibration condition monitoring. Therefore, how to apply machine-learning approaches for decision making with higher accuracy in vibration condition is always a hot issue.

Typically, electronically measured raw signals are transformed in such a way that levels of these quantities describe the condition of a given machine. Raw vibration signals are transformed using analysing techniques such as Fast Fourier Transform (FFT), Short Time Fourier Transform (STFT), empirical mode decomposition (EMD), and Wigner-Ville distribution (WVD). All these methods have been widely applied to extract patterns in either time domain or frequency domain from raw vibration data, which can be subsequently leveraged for fault identification and classification. However, for degradation assessment in mechanical equipment, it is significant to represent the vibration data in both time and frequency domains to track and map the changes of degrading. Although FFT based technologies are useful for fault classification and identification, they are usually not suitable for non-stationary signals. To analyse data in the time-frequency domain, WVD and STFT were once the most popular methods for non-stationary signal. However, WVD suffers from interference terms appearing in decomposition, while STFT cannot provide ideal time and frequency resolution simultaneously since it applies constant resolution at all frequencies. In addition, no orthogonal bases exist for SFT that can be leveraged to implement a both fast and effective STFT algorithm [Okumura, 2011; Vachtsevanos et al., 2006].

Under this background, nowadays, wavelet transform based technologies such as WPD has been increasingly applied in many cases due to the great capabilities in both time and frequency domains. With the help of wavelet transform, the analysis of non-stationary signals is achievable as well as detecting transient feature components as other methods were inept to perform since wavelet can concurrently impart time and frequency structures. For this reason, in this chapter, we leverage WPD to represent the working condition of mechanical equipment through features in both time and frequency domain from vibration signals, and subsequently apply extracted information for fault classification and degradation assessment.

6.3 WPD

WPD is a very useful tool to analyze vibration signals. In numerical analysis, the essence of WPD is a wavelet transform where the discrete-time signal is parsed through more filters than the discrete wavelet transform, which can provide a multi-level time-frequency decomposition of signals [Y. Zhang et al., 2016]. It is extended from the wavelet decomposition (WD) and includes multiple bases and different basis, which can result in different classification performance and cover the shortage of fixed time–frequency decomposition in Discrete Wavelet Transform (DWT) [Xue et al., 2003].

In DWT, the original signal will first pass through two complementary filters and emerges as approximation coefficients and detail coefficients, which includes the low frequency and high frequency information about the original signal respectively. The approximation coefficient will further split into a second-level approximation coefficients and detail coefficients. This process may repeat according to the number of decomposition layers. Figure 6.1 shows the 3-layer structure of signal based on DWT, where approximation coefficients and detail coefficients are labelled as A and D respectively.

However, WPD decomposes the detail and approximation coefficients simultaneously. Therefore, WPT can construct a complete wavelet packet tree with the same frequency bandwidths in each resolution. WPD can lead to a complete wavelet packet tree as shown in Figure 6.2. A wavelet packet is a function with three parameters, *i*, *j* and *k*, which are the modulation, scale and translation parameters respectively.

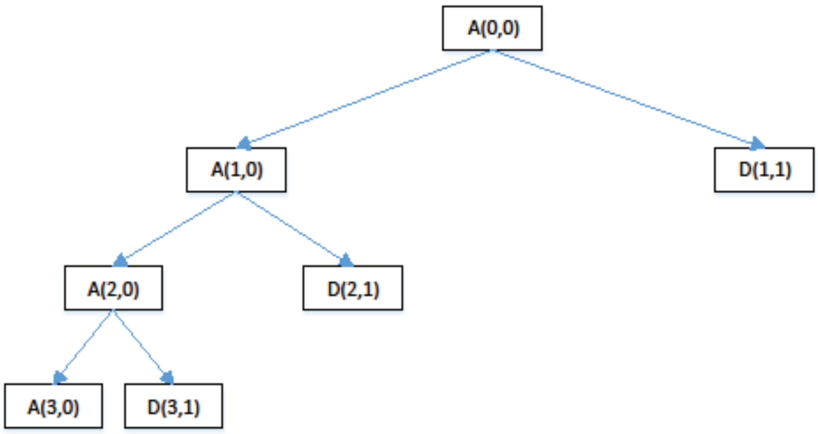

Figure 6.1 3-layer structure of DWT

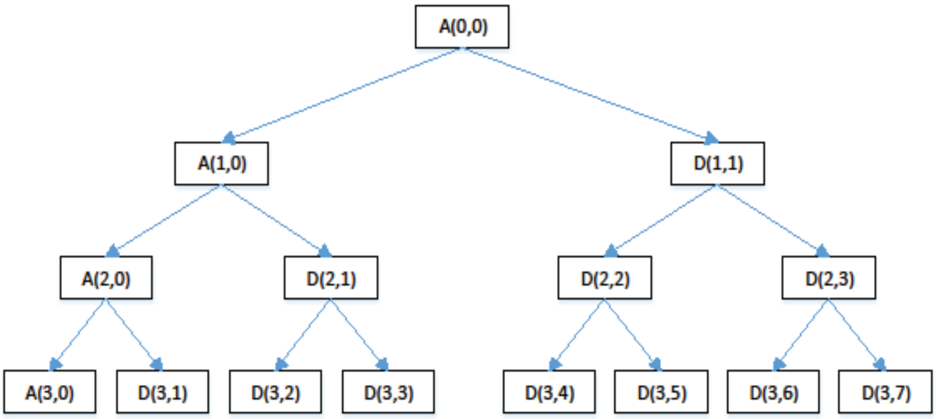

Figure 6.2 Wavelet packet decomposition

According to the space partition shown in Figure 6.2, we label W_j^i to represent the i th subspace of wavelet packet at the j th scale, where $W_{j,k}^n(t) = 2^{-\frac{j}{2}}\omega^n(2^{-j}t - k)$, k is the shift factor and $k \in Z$ [Ting et al., 2008]. It satisfies with equation (6.1) and (6.2).

$$\omega_{j,0}^n(t) = \sum_k h_0(k)\,\omega_{j-1,k}^i \qquad (n \text{ is odd}) \qquad (6.1)$$

$$\omega_{j,0}^n(t) = \sum_k h_1(k)\,\omega_{j-1,k}^i \qquad (n \text{ is even}) \qquad (6.2)$$

Where $j,k \in Z, n = 0,1,2,\dots,2^i - 1, h_0(k), h_1(k)$ are low-pass and high-pass filters of wavelet packet. Then the original signal $f(t)$ can be represented according to j level WPD as equation (6.3) and (6.4). The wavelet packet component $f_j^i(t)$ can be obtained through a linear combination of wavelet packet function $W_{j,k}^n(t)$ and wavelet packet coefficients $c_{j,k}^i$.

$$f(t) = \sum_{i=1}^{2j} f_j^i(t) \qquad (6.3)$$

$$f_j^i(t) = \sum_{-\infty}^{+\infty} c_{j,k}^i(t) W_{j,k}^i(t) \tag{6.4}$$

Different types of wavelet functions may cause different time-frequency structures, in this chapter, Daubechies 4 (DB4) wavelet function has been chosen due to the good performance in estimations of the local properties of signals like breakdown points [Ferreira and Borges, 2003], and the ability to derive a set of conventional and energy based features from signals [Murugappan et al., 2010]. During the test, WPD is applied to extract the standard deviations of coefficients and the percentage of energy corresponding to the approximation and details to represent the working condition.

6.4 Set up and data collection

During the experiment, a Bently Nevada Rotor Kit RK3 is used to simulate the real working condition of rotary machinery. A sleeve-bearing house is equipped with three accelerometers of Kistler 8702B100, mounted in X, Y, Z three directions, to measure the vibration signals from the test rig, as shown in Figure 6.3. The sampling frequency is 4096 HZ and the maximum revolving speed of the rotor kit during the experiment is 4000 rpm. The bearing block is tightened down to the foundation and can be loosened during the experiment. Rub generator and mass adjustable load can be modulated to simulate types of failures. The vibration monitoring refers to a zero position of the test rig. In this position, signals from the accelerometers are recorded and stored.

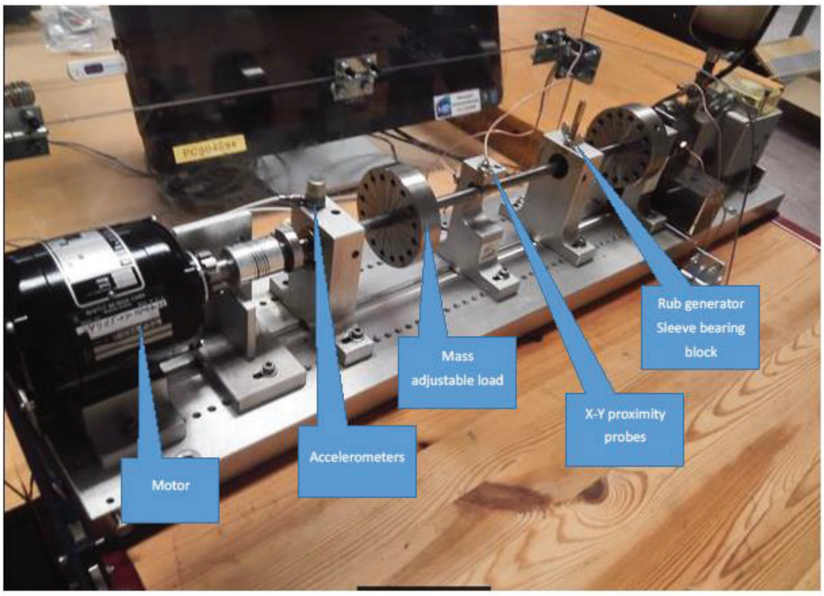

Figure 6.3 Bently Nevada rotor kit

During the experiment, we injected three types of failures, bearing looseness by loosening bearing housing, main spindle friction by applying rub generator force to the rotor kit axel, and load imbalance through adding weights, as shown in Figure 6.4, to the flywheel. For each type of failure, the vibration signals will be measured through use of accelerometers at different failure degradation and rotating speed by means of proximity sensors and hand held tachometer for control. Exact measurement of rpm is of utmost importance as vibration relates to this frequency by whole or half numbers. Figure 6.5 shows the transform from raw vibration signals to the wavelet coefficient-based and energy-based features.

Figure 6.4 Weights on flywheel

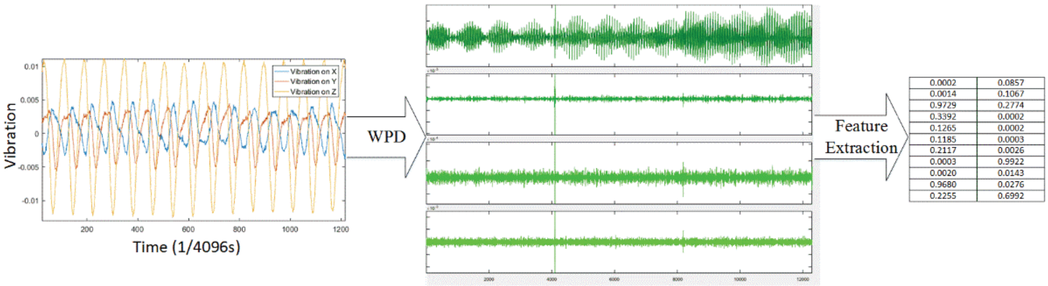

Figure 6.5 Vibration signals and the wavelet coefficients and energy features

After extracting the wavelet coefficients and energy features from vibration signals, a common dilemma when analysing vibration data from mechanical equipment is to determine the vibration level acceptance criteria. It is also a challenge when using WPD for fault classification and degradation assessment in mechanical equipment. In order to solve this challenge, deep neural network with BP will be introduced and applied to analyse vibration data for fault classification and degradation assessment in next section.

6.5 Deep neural network with BP

Deep neural network with BP is a type of neural networks with multiple hidden layers, trained through backpropagation procedure. This kind of networks is one of the most original deep learning approaches. Although all types of deep learning networks can be called as "deep neural networks", here we labelled fully connected deep neural network with back propagation as DNN to distinguish it from other types of deep learning architectures. As we discussed above, researchers found it was difficult to train a multilayer neural network which is constructed through stacking number of hidden layers directly in practice in the earliest days. However, due to the dramatic advance of current computational power and development of training algorithms, modern DNN can be considered as a powerful approach for modelling high complexities.

DNN grows from one of the most widely used neural network model, BPNN, with higher complexity through increasing the number of hidden layers. BPNN, firstly proposed by Rumelhart and McCelland in 1985 [Rumelhart et al., 1985], is a multilayer feed-forward network usually containing three layers, the input layer, the hidden layer and the output layer. Neurons, which are setup in each layer, are fully connected between different layers. The number of neurons in the input and output layer equals to the dimension of the inputs and outputs, respectively. The number of neurons in the hidden layer is adjustable, as well as the neuron amount. Each connection between neurons represents an activation function converting the neurons weight to corresponding output. This is the reason why these neurons are also called as nodes. A BPNN can define a function $f: X \rightarrow Y$ to map the input dataset X to output dataset Y.

DNN can be considered as an evolutionary type of BPNN with multiple hidden layers (at least three hidden layers) and largely increasing complexity. A DNN, as shown in Figure 6.6, consists of one input layer, numbers of hidden layers, and an output layer, forming the topology of the net. The input layer matches the feature space, so that there are as many input neurons as predictors. The output layer is either a classification or regression layer to match the output space. All layers are composed of neurons, which is also the basic units of such a model like BPNN. It is also called as deep feedforward networks neural networks, since it follows the classical feedforward architecture, where each neuron in the previous layer L is fully connected with all neurons in the subsequent layer L+1 via directed edges, each representing a certain weight. Also, each non-output layer of the net has a bias unit, serving as an activation threshold for the neurons in the subsequent layer. As such, each neuron receives a weighted combination of all the outputs of the neurons in the previous layer as input [Krauss et al., 2017]. This type of models is called feedforward because information flows through the function being evaluated from x, through the intermediate computations used to define f , and finally to the output y. There are no feedback connections in which outputs of the model are fed back into itself [I. Goodfellow et al., 2016]. For a DNN, the number of hidden layers in the network is adjustable, just like the number of nodes is adjustable in a BPNN.

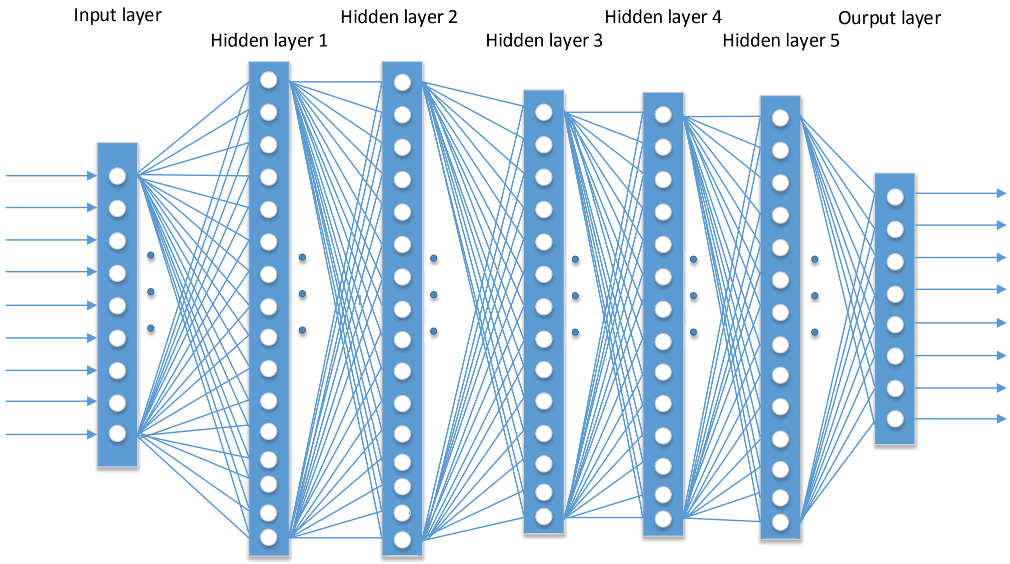

Figure 6.6 Architecture of DNN

In BPNN, neurons with sigmoid activation functions are trained through gradient descent learning approach, since the structure of BPNN is relatively simple and the complexity of the model is low. As to DNN, it is almost impossible or extremely hard to train the network through this method because of the gradients vanishing. As we discussed in Chapter 2, although the computational power of today's computers is million times the computational power of the early 1990s', which allows for propagating errors a few layers further down within reasonable time. However, it does not really overcome the problem in a fundamental way. As reported by Nielsen [2015], when use the learning method of conventional BPNN to train a DNN, it is discovered that the different layers in the network are learning at vastly different speeds. In particular, when later layers in the network are learning well, early layers often get stuck during training, learning almost nothing at all.

To deal with this fundamental problem, modern DNN replace the sigmoid activation functions with alternative functions such as Tanh and Maxout and apply advanced training procedure like momentum-based stochastic gradient descent [Sutskever et al., 2013] and parallelizing stochastic gradient descent [Recht et al., 2011]. It is typically found in sigmoid networks that gradients vanish exponentially quickly in earlier layers, which greatly slows down the learning procedure. In addition, Glorot and Bengio [2010] have provided the evidence that the use of sigmoid activation functions would cause the activations in the final hidden layer to saturate near 0 early in training, substantially slowing down learning. Several alternative activation functions currently employed for DNN, along with their formulas, are listed in table 6.1 [Candel et al., 2015]. x_i and $w_i$are marked as the input values of the firing neurons and their weights, respectively. α represents the weighted combination $\alpha = \sum_i x_i w_i + b$.

Table 6.1 Activation functions for DNN

Function	**Formula**	**Range**
Tanh	$f(\alpha) = \frac{e^{\alpha} - e^{-\alpha}}{e^{\alpha} + e^{-\alpha}}$	$f(\alpha) \in [-1,1]$
Maxout	$f(\alpha_1, \alpha_2) = \max(\alpha_1, \alpha_2)$	$f(\alpha) \in R$
Rectified Linear	$f(\alpha) = \max(0, \alpha)$	$f(\alpha) \in R_+$

The hyperbolic tangent (tanh) function is a typical choice for DNN. This function is defined as the ratio between the hyperbolic sine and the cosine functions or expanded as the ratio of the half-difference and half-sum of two exponential functions in the points α and $-\alpha$ as Equation (6.5). The symmetry around 0 allows the training algorithm to converge faster.

$$f(\alpha) = \frac{\sinh(\alpha)}{\cosh(\alpha)} = \frac{e^{\alpha} - e^{-\alpha}}{e^{\alpha} + e^{-\alpha}} \tag{6.5}$$

The rectified linear activation function has demonstrated high performance on image recognition tasks and is a more biologically accurate model of neuron activations [LeCun et al., 2012]. Maxout is a generalization of the rectified linear activation, where each neuron picks the largest output of k separate channels. And each channel has its own weights and bias values. Maxout activation functions work particularly well with dropout (Drop is a technique that can be applied to deterministic feedforward architectures that predict an output given input vector) [I. J. Goodfellow et al., 2013].

Due to the strong capability of self-study, BPNN could overcome the bottle-neck of knowledge obtain in many situations. Actually, BPNN was one of the most widely used neural network and a popular approach for failure detection, classification, and faults prediction in various research fields such as motor bearing diagnosis [D.-M. Yang et al., 2002], urban water mains [Jafar et al., 2010], and transformer [Sun et al., 2007]. The process of failure detection or prediction actually is a pattern discriminating fundamentally, which can be interpreted as a process to map the character space X to faults space Y, The mapping process for fault diagnosis or prognosis is usually highly non-linear, which could be simulated by a multilayer back-propagation neural network. DNN as an evolutionary version of BPNN is also widely applied in fault diagnosis and prognosis to pursue higher accuracy for faults detection, classification, or prediction. Compared with other conventional machine learning algorithms, the superiority of DNN is in degradation mapping, and failures identification, when enough history data could be obtained, and the complexity of target issue is relatively high. In next section, we also provides an experiment of fault classification and degradation assessment in rotary machinery, in which DNN outperforms other data-driven methods and demonstrates its advantages in degradation mapping. DNN models applied here are constructed through five hidden hyperbolic tangent layers with 50 * 50 * 32 * 32 * 32 nodes, and trained through parallelizing stochastic gradient descent.

6.6 Faults classification and degradation assessment

6.6.1 Numerical result

During the experiment, 10216 samples of data have been collected to establish and test the data-driven models for failure classification and degradation assessment, respectively. Among those samples, 6817 samples are collected for fault classification and 3399 samples are applied for degradation assessment, respectively. Table 6.2 illustrates the data composition of collected samples for fault classification. The samples for training and testing during the experiment are selected stochastically.

Table 6.2 Data composition of collected samples for fault classification

6817 samples collected in different rotating speed for fault classification	1608 samples collected in normal working condition (failure type 0)	1376 samples for training (stochastic)
		232 samples for testing (stochastic)
	1703 samples collected in failure type 1: Friction on main spindle	1551 samples for training (stochastic)
		152 samples for testing (stochastic)
	1782 samples collected in failure type 2: Bearing looseness	1555 samples for training (stochastic)
		227 samples for testing (stochastic)
	1724 samples collected in failure type 3: Load imbalance	1549 samples for training (stochastic)
		175 samples for testing (stochastic)

The training and testing process have been run 5 times during the experiment with stochastic selection for training and testing samples. Figure 6.7 shows one of the classification results from DNN when misjudgement occurred, where the correct classification rate is 99.87% in this test. Through observation, the proposed method can ideally classify the failures through coefficients and energy based features for rotating machinery with tiny fluctuations. There are no significant divergence among the results.

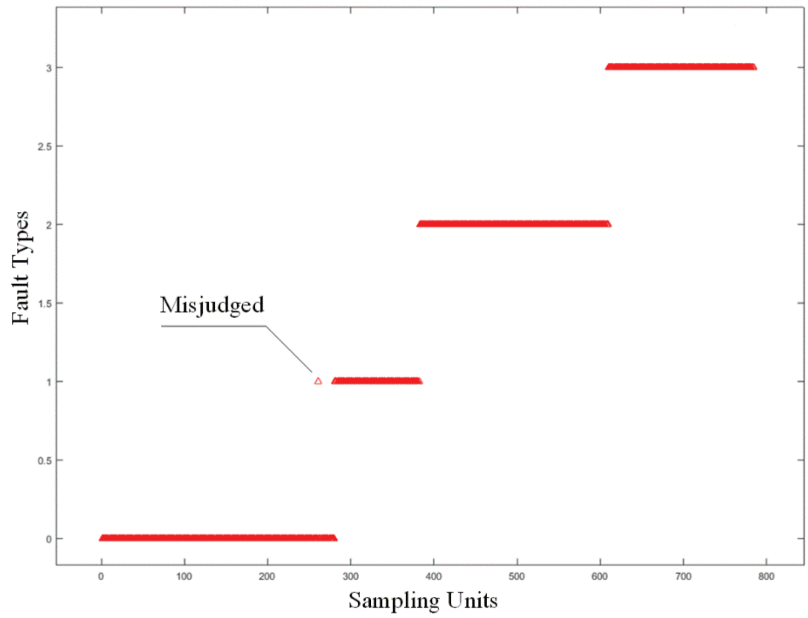

Figure 6.7 Classification result from DNN

During degradation assessment, DNN is also leveraged to estimate the degradation of imbalance on spindle. In this stage, 3399 samples are collected under different speed with changeable weight on the flywheel to simulate the degradation of load imbalance. The degradation is adjusted by changing weight on the mass adjustable load, which raised from 0.25 to 8 grams (the values has been multiplied by 100 for calculation and visualization). The weight was used to represent the degradation of load imbalance.

Therefore, the inputs of the data-driven models are the coefficient and energy based features extracted from WPD, and outputs are the estimated weight on mass adjustable load, which would cause load imbalance. The training and testing processes have been run for 5 times with stochastic selection for training and testing samples (About 10 percent of collected samples are used as testing samples). According to the numerical result, there are no significant divergence among the results. Figure 6.8 shows part of the assessment results based on the proposed method. In that test, the number of training and testing samples are 3089 and 310, respectively.

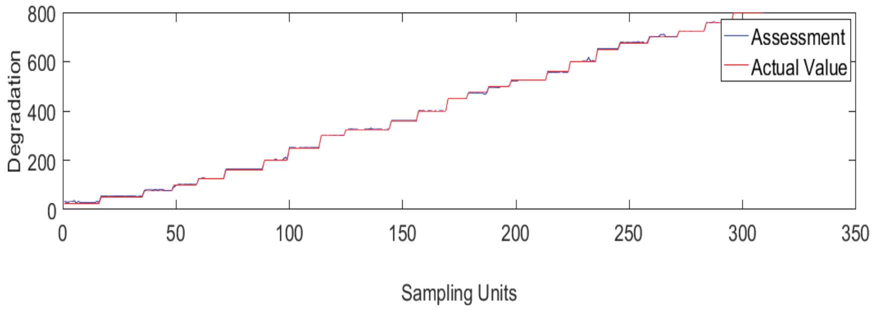

Figure 6.8 Degradation assessment of DNN

6.6.2 Comparison with conventional methods

In order to provide a comprehensive comparison of different types of data driven models, we also applied other popular and widely applied methods such as SVM [Ramesh Babu and Jagan Mohan, 2017], DBN [Z. Zhang and Zhao, 2017], KNNC [Ha et al., 2017], and BPNN [Asuhaimi Mohd Zin et al., 2015]. We also run 5 times for those 4 types of data driven models with stochastic selection for the training and testing process to verify their performance in our case. Figure 6.9 demonstrates the mean correct classification rates (MCCR) of all the applied data-driven models together with DNN. During the test, all these five types of data-driven models can classify the failures through coefficients and energy based features for mechanical equipment in a great performance and with tiny fluctuations. Since the performances of all the methods are acceptable during the test, the numerical results shows that all applied methods have the ability for failure classification based on the information extracted from vibration signals in the rotating equipment.

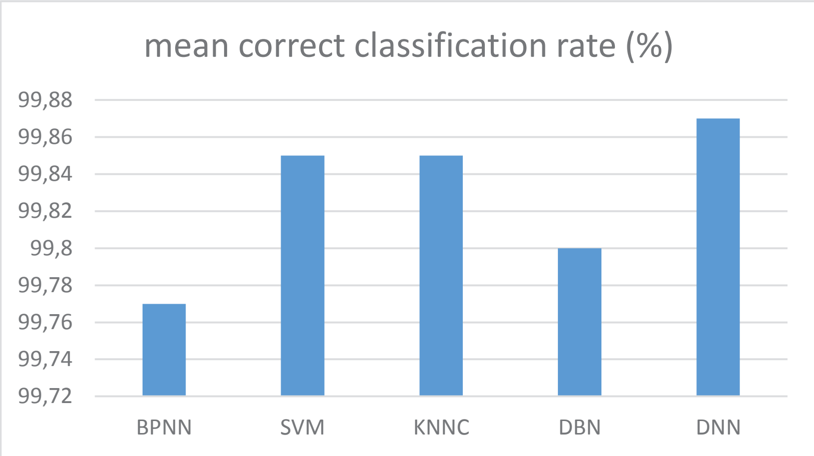

Figure 6.9 Mean correct classification rates in fault classification

As to degradation assessment, the samples for testing and training are fixed this time to offer a more visualized comparison. The training and testing processes have also been run for 5 times with fixed training and testing samples (3089 training samples and 310 testing samples trained in DNN). Figure 6.10 - 6.13, show the assessment results based on those methods, respectively.

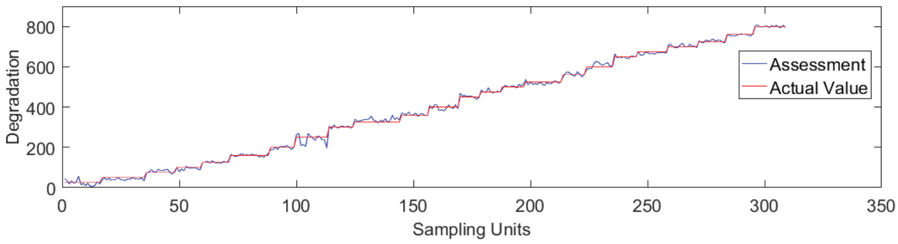

Figure 6.10 Degradation assessment of BPNN

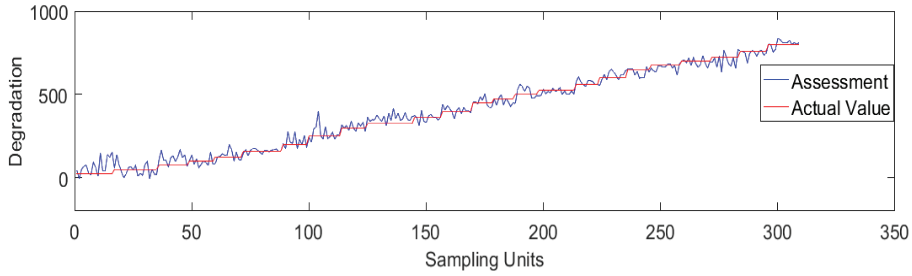

Figure 6.11 Degradation assessment of SVM

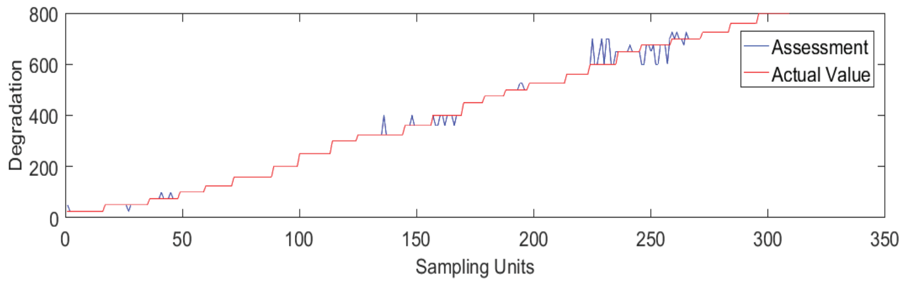

Figure 6.12 Degradation assessment of KNNC

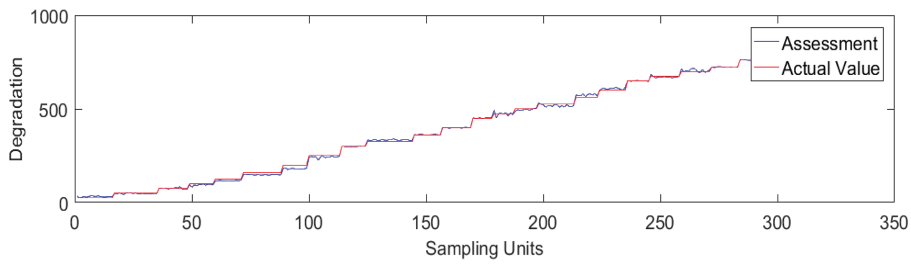

Figure 6.13 Degradation assessment of DBN

According to the assessment results, it can be figured out that DNN shares the best performance to map the actual degradation of imbalance for the equipment during the experiment. The assessed degradation through DNN almost completely coincide with the actual values of weight attached on the flywheel. Although the assessment errors from DBN and BPNN are much higher than DNN and kept fluctuating during the test, the overall assessment results still prove that they have the ability to map the degradation in certain degree. As to SVM and KNNC, both of them demonstrate their capacity to deal with fault classification, however, the performances for degradation assessment during the experiment are unacceptable. To obtain more detail information from the numerical results, we also run 5 times for those 4 types of data driven models with stochastic selection. The mean square error (MSE) and largest error (LE) are calculated and recorded to acquire a comprehensive comparison for the stability and overall performance during the experiment. As shown in Table 6.3, detail information about the numerical results is listed, including mean square error (MSE) and largest error (LE) along with the structures of the five data-driven models applied for degradation assessment.

Table 6.3 Numerical results of degradation assessment

Model	Structure	MSE($\frac{gram^2}{10^4}$)	LE($\frac{gram}{10^2}$)
BPNN	50 nodes in the hidden layer	112,8937	51,2
SVM	Gaussian Kernel Function	1230,3144	145,3
KNNC	Standard Model	307,0388	97,5
DBN	5 layers, 50 * 50 * 32*32*32	66,2543	23,99
DNN	5 layers, 50 * 50 * 32 * 32*32	9,6274	15,99

6.6.3 Discussion

The DNN leveraged during the experiment is constructed with five hidden layers with hyperbolic tangent function and trained through parallelizing stochastic gradient descent. The applied DBN model is stacked through stacking five RBM layers with Bernoulli functions. The structures applied during the experiment are selected according to empirical knowledge through input dimensions, training time, and complexity of the issue. Simultaneously, the models of SVM and KNNC are with Gaussian kernel function and standardize model since they are the most popular and widely applied ones. The numerical results show that DNN with 5 hidden layers has the best performance for degradation assessment with 9,6274 $gram^2/10^4$ MSE and 15,99 $gram/10^2$ LE, which greatly outperformed others.

During the test, SVM shows good performance in fault classification with 99.85% mean correct classification rate, but the ability to map the degradation is the worst of the five tested models. The application of SVM for degradation assessment is also well presented in [Soualhi et al., 2015]. In that paper, the authors firstly defined three degradation states represented by three classes according to the degradation of bearing. Then SVM is leveraged to detect the degradation states of bearings with good performance. The main target in that paper lies on bearing health monitoring instead of accurately degradation assessment. According to our experiment, we consider SVM is good at classifying or recognizing failure patterns from fault information instead of mapping degradation directly. Similarly, KNNC is also a perfect tool for identification of the degradation mode or level. As well described in [Baraldi et al., 2016], in which the authors successfully applied KNNC as a diagnostic approach for the identification and characterization of defeats in automotive bearings. In that paper, the author also argued that though the performance of KNNC has been reported to be less satisfactory in some applications than that of other popular data driven models, the classifiers of KNNC have advances in simplicity and low computational requirements.

Although not as good as DNN, BPNN and DBN also show the ability to deal with the issues about degradation assessment. In addition, the performances of DBN and BPNN are not quite stable according to the MSE during the experiment. As reported in [C. Zhang et al., 2017], DBN is capable of extracting a hierarchy of features, where features at higher network levels are usually more relevant to the ultimate task. The authors in that paper evaluate the performance of DBN on several prognostic benchmarking data sets and prove its superiority in estimating remaining useful life. In addition, we also investigated the performance of DBN in Chapter 6, in which DBN proved its superiority when the target is beyond history data.

Based on these considerations, each data driven methods may have their own advantages in certain domains or with prerequisite. Actually, this is the reason why it is significant to sort the superiorities of all widely applied deep learning algorithms for predictive maintenance in Chapter 2. During the test, DNN proves its perfect ability to identify and classify failures for rotating equipment. Simultaneously, the performance of degradation assessment greatly outperformed the methods proposed in literature.

6.7 SAE-LSTM anomaly detection

In the last section, the superiority of DNN-based degradation assessment has been demonstrated through numerical result. However, in many practical applications, one may face the dilemma that the history data is collected and recoded unlabelled, let alone classified. To solve this challenge, a SAE-LSTM anomaly detection method is proposed in this section to identify the anomaly condition in an unsurprised learning environment.

6.7.1 SAE-based representation learning for multiple features sequence

When the history data is collected without labels (These labels usually can be used to represent the working condition in a surprised learning manner), an alternative method is to track the changes in multiple features sequence with time-series to identify the anomaly. To prevent the inputs from explosion, SAE-based representation learning is leveraged to reduce the number of features extracted from raw data, and reconstruct the multiple features sequence as shown in Figure 6.14.

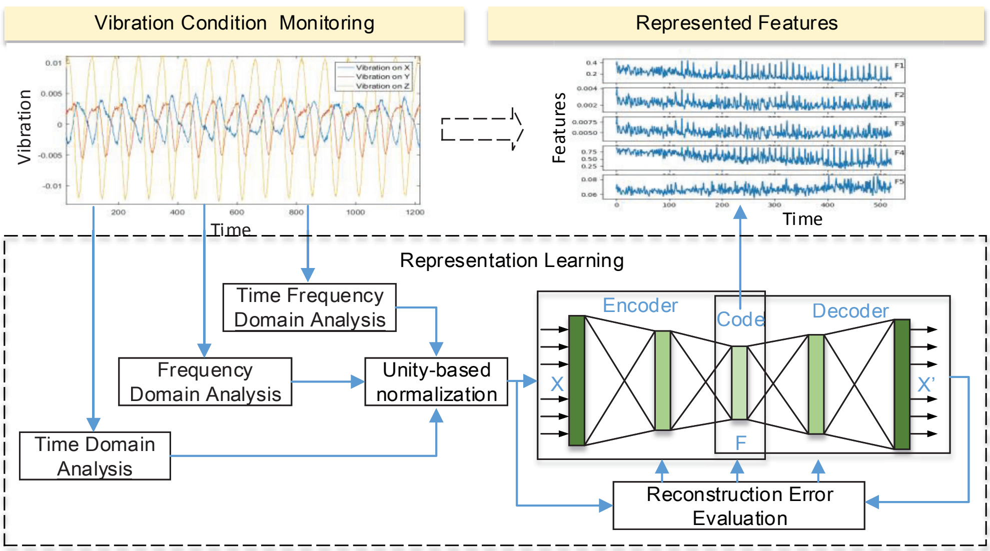

Figure 6.14 Process of SAE-based representation learning for multiple features sequence

6.7.1.1 Feature normalization

During the experiment, vibration signals in both normal and anomalous condition are collected through 4096HZ sampling frequency. Each sampling units have been divided into 10 parts with the same length for time series-based detection. The coefficient-based and energy-based features of each parts were extracted through WPD. Table 6.4 shows part of the energy-based features through DB4 wavelet transform. Eax denotes the percentage of energy corresponding to the approximation in x direction and Edx means the vector containing the percentages of energy corresponding to the details at each layer in x direction.

Table 6.4 Part of the energy-based features through DB4 wavelet transform

No.	Eax	Edx(1)	Edx(2)	Edx(3)	Edx(4)
1	92,52748	0,49676	0,206692	0,216633	6,552434
2	92,6626	0,484495	0,196127	0,206738	6,450044
3	92,65699	0,512164	0,210155	0,212048	6,408645
4	92,66757	0,462788	0,192835	0,243686	6,433121
5	92,19917	0,510604	0,204128	0,238897	6,847205
6	92,43177	0,49443	0,196106	0,221158	6,656534
7	92,43952	0,488822	0,212711	0,208403	6,650545
8	92,54461	0,458843	0,207961	0,197875	6,590716
9	92,53294	0,480579	0,2039	0,194012	6,588572
10	93,06981	0,440835	0,168283	0,175664	6,145403
11	93,07948	0,470262	0,177757	0,195515	6,076985
12	92,62898	0,512523	0,213712	0,194187	6,450598
13	92,76121	0,487038	0,203332	0,19106	6,357365
14	92,6972	0,482968	0,193362	0,202178	6,424294
15	92,31745	0,513153	0,207133	0,22472	6,737547
16	92,48647	0,498813	0,201542	0,213873	6,599307
17	93,74498	0,421325	0,170193	0,168667	5,49483
18	93,07432	0,438584	0,176644	0,195416	6,115036
19	92,83945	0,486567	0,193284	0,187082	6,29362
20	92,00818	0,551392	0,233748	0,220501	6,986177

To adjust values measured on different scales to a notionally common scale, unity-based normalization has been applied to normalize the inputs F'_{ij} for SAE-based representation learning.

$$F'_{ij} = \frac{F_{ij} - F_i^{min}}{F_i^{max} - F_i^{min}} \tag{6.6}$$

Where F_{ij} denotes the i^{th} feature in j^{th} samples, F_i^{min} and F_i^{max} represent the minimum and maximum values of the i^{th} feature in database, respectively. Figure 6.15 shows part of energy-based features after normalization, which will be used as inputs for SAE-based dimension reduction (Anomaly sampling units are collected when failures are injected, but the labels will be used in validation only). Visually, after normalization, data collected in anomaly still keep certain divergence from normal condition, though we cannot catch the rules directly in this step. It should be noticed that, during the training process, it is supposed that we only have the data in normal condition. Data in anomaly is only collected to test and validate the proposed method.

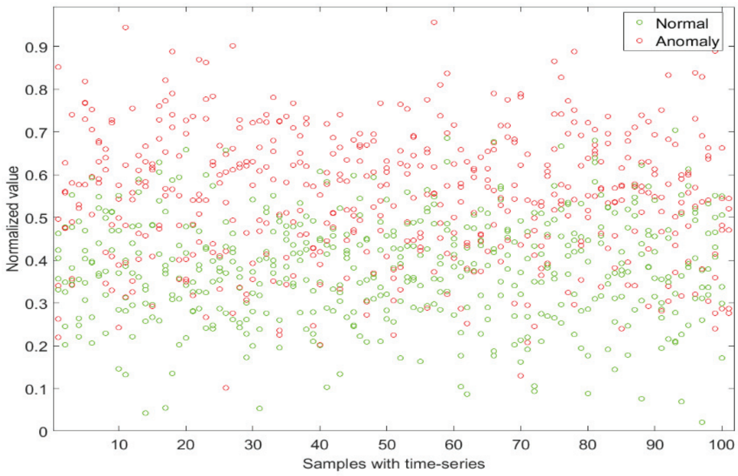

Figure 6.15 Part of energy-based features after normalization

6.7.1.2 Feature representation with SAE

After feature normalization, sparse autoencoders have been leveraged to construct the deep neural network for representation learning. SAE is first proposed in 2007 [Bengio et al., 2007; Poultney et al., 2007]. It is a special type of deep neural networks created through stacking multiple autoencoder layers. The architecture of the deep neural network is pre-trained through single autoencoder layer by layer [Bengio et al., 2013]. The output of SAE is the data input itself, which is leveraged for learning efficient encoding or dimensionality reduction for a set of data. More specifically, it is a nonlinear feature extraction method involving no class labels; hence generative. An autoencoder uses three or more layers in the neural network, and when the number of hidden layers is greater than one, the autoencoder is considered to be deep [Deng, 2012].

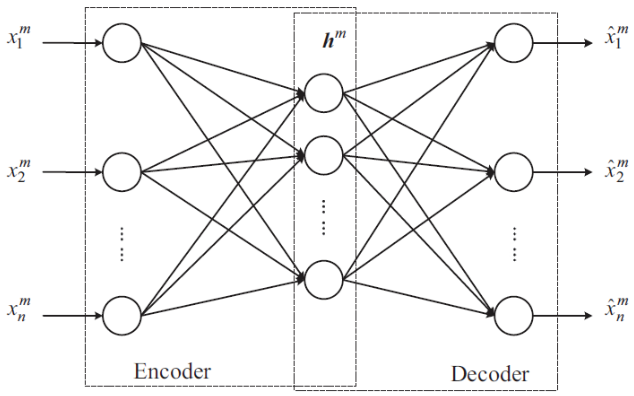

Figure 6.16 Architecture of an autoencoder

As depicted in Figure 6.16, the input layer and hidden layer construct the encoder network, which transforms the input data from a high-dimensional space into codes as a low-dimensional space and the decoder network, which consists of the hidden layers and output layer, reconstructs the inputs from the corresponding codes. The encoder network is explicitly defined as an encoding function denoted by f_θ, which is also called as the encoder [Bengio et al., 2013]. For each input signal $\boldsymbol{x}^m$ from a dataset $\{\boldsymbol{x}^m\}_{m=1}^M$, we label h^m as the obtained encode vector:

$$\boldsymbol{h}^m = f_{\boldsymbol{\theta}}(\boldsymbol{x}^m) \tag{6.7}$$

The decoder network is defined as a reconstruction function denoted by $g_{\theta'}$, namely the decoder. It maps $\boldsymbol{h}^m$ from the low-dimensional space back into the high-dimensional space, producing a reconstruction as Equation (6.8):

$$\widehat{\boldsymbol{x}}^m = g_{\boldsymbol{\theta}'}(\boldsymbol{h}^m) \tag{6.8}$$

The parameter sets of the encoder and decoder are learned simultaneously on the task of reconstructing as well as possible the original input, attempting to incur the lowest possible reconstruction error $L(\boldsymbol{x}, \widehat{\boldsymbol{x}})$ over the M training examples. $L(\boldsymbol{x}, \widehat{\boldsymbol{x}})$is a loss function that measures the discrepancy between $\boldsymbol{x}$ and $\widehat{\boldsymbol{x}}$ [Bengio et al., 2013].

In summary, the autoencoder training aims to find the parameter sets $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$ minimizing reconstruction error, which can be depicted as Equation (6.9)

$$\varphi_{AE}(\boldsymbol{\theta}, \boldsymbol{\theta}') = \frac{1}{M}\sum_{m=1}^{M} L(\boldsymbol{x}^m, g_{\boldsymbol{\theta}'}(f_{\boldsymbol{\theta}}(\boldsymbol{x}^m))) \tag{6.9}$$

A deep neural network could be constructed by stacking multiple autoencoder layers with a final classification or regression layer on top. Stacking multiple autoencoder layers together allows the network to learn higher order features, where each successive layer represents additional complexity within the input data [Galloway et al., 2016]. Each hidden layer in SAE is pre-trained through learning multiple nonlinear transformation of the inputs indecently. With the strong ability of self-learning, SAE could capture the main variations, discover the discriminative information, and represent the features from the raw data in an unsupervised manner [Erhan et al., 2010]. For predictive maintenance, representations of working condition with lower-dimension can improve performance in many situations such as fault classification and detection, especially when the input data is industrial big and row data. As discussed in Chapter 2, many practical applications have shown that SAE has the ability to automatically mine the important information from the frequency spectra according to the diagnosis issues. With the code vector of the previous trained autoencoder as input for training the next autoencoder, SAE could recognize the characteristics and effectively discover the discriminative information of these signals, and subsequently represent mechanical health conditions. The SAE constructed during the test has three hidden layers trained through L2 regularization. The original data includes 33 features in both time and frequency domains. After representation learning, the features are transformed into a multiple features sequence with time series, as shown in Figure 6.17

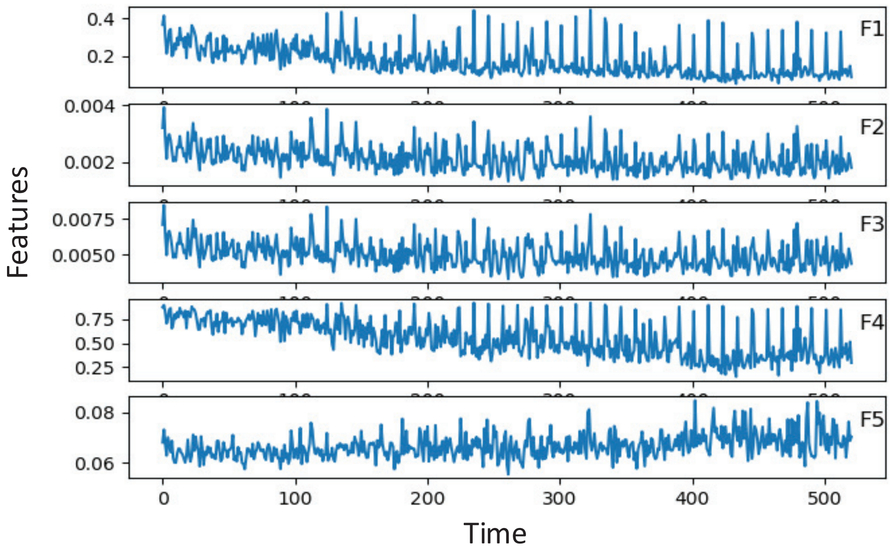

Figure 6.17 Features after representation learning

6.7.2 LSTM-based anomaly identification with time series

6.7.2.1 Architecture of LSTM

As we discussed above, to solve the fundamental problem of gradient vanishing, an alternative method is to employ special architectures unaffected by it. LSTM is the most typical model in this type of deep learning architectures, which could avoid the fundamental problem of gradients vanishing through special architectures. LSTM neural network is a type of recurrent neural network proposed in 1997 to address the problem of insufficient, decaying error backflow in RNN training [Hochreiter and Schmidhuber, 1997]. The basic idea of LSTM is simple: In a LSTM neural network, memory cells are employed as independent activation functions and identity functions with fixed weights, which are connected to themselves. Due to fixed weight, errors back-propagated through a memory cell cannot vanish or explode but stay as they are [Schmidhuber, 2015]. The weight matrixes in conventional RNNs are also trained via backpropagation through time series like the training process of normal neural network. Therefore, the gradients vanishing problem also happens in RNN while the complexity of the network increases, which means traditional RNN do not have the ability to discover information or capture dependencies hidden in long-term time series. In this background, LSTM was proposed to prevent back propagated errors from gradients vanishing or exploding in RNN to deal with issues about long-term dependencies. The core idea behind the LSTM architecture is a memory cell, which can maintain its state over time, and non-linear gating units regulating the information flow into and out of the cell [Greff et al., 2016]. Compared with traditional RNN, LSTM neural network leverages memory cells with forget gates instead of traditional neurons to establish connections between inputs and outputs. These adopted forget gates can effectively control the utilization of information in the cell states, and enable LSTM the capability to capture nonlinear dynamics in time series sensory data and learn effective representation of machine [Zhao et al., 2017]. LSTM applies four special and interacting neural network layers, layer α, β, γ , o , instead of a single layer as in a standard RNN [Liao and Ahn, 2016], as shown in Figure 6.18. The first layer α is a sigmoid layer also called as forget gate layer, which returns a value between 0 and 1 in the previous cell state C_{t-1} , while 0 means no information pass and 1 means all information pass. The equation of the first layer can be denoted as Equation (6.10).

$$\alpha_t = \sigma(W_\alpha \cdot [h_{t-1}, x_t] + b_\alpha) \tag{6.10}$$

Where σ is the sigmoid function, W_α is the weight of layer α, [ ] denotes the concatenate operation, x_t is the input x and time t, h_t is the output with respect to x_t, W_α , W_β , W_γ , W_o are the weights and b_α , b_β , b_γ, b_o are the biases of the layer α, β, γ, o, respectively. The second layer β is called as input gate layer, which is applied to decide which value shall be updated, denoted as Equation (6.11)

$$\beta_t = \sigma\left(W_\beta \cdot [h_{t-1}, x_t] + b_\beta\right) \tag{6.11}$$

Next, a tanh layer γ updates the values to be stored using:

$$\gamma_t = tanh\left(W_\gamma \cdot [h_{t-1}, x_t] + b_\gamma\right) \tag{6.12}$$

Where $tanh$ is the hyperbolic tangent function.

Then, we can update the previous state C_{t-1} to the current state C_t by Equation (6.13)

$$C_t = \alpha_t \cdot C_{t-1} + \beta_t \cdot \gamma_t \tag{6.13}$$

The final layer is also a sigmoid function layer, which determines what parts of the cell state will be the output, as denoted by Equation (6.14)

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{6.14}$$

Then, the cell state go through tanh function and form the final output as Equation (6.15)

$$h_t = o_t tanh(C_t) \tag{6.15}$$

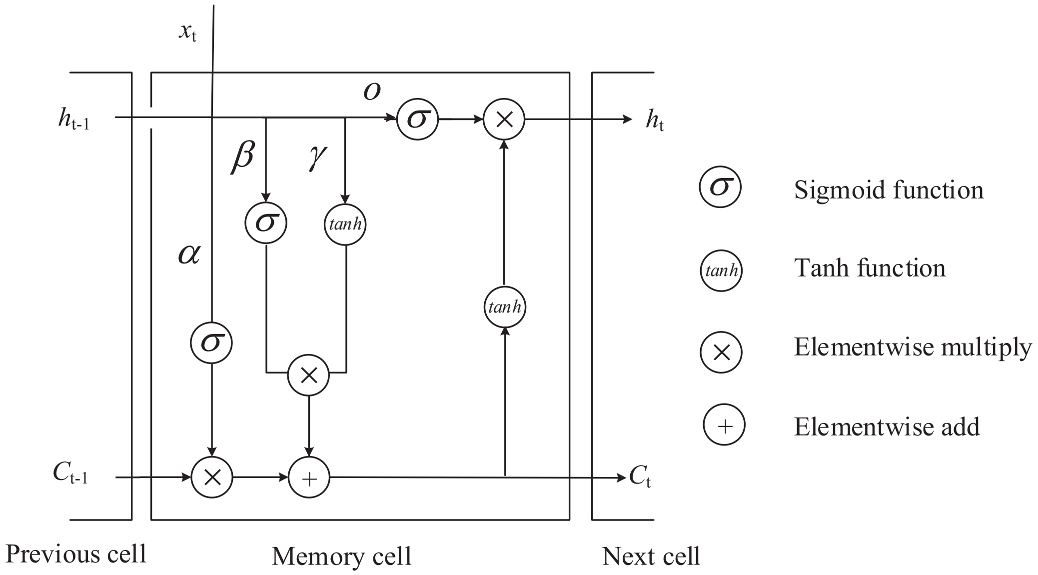

Figure 6.18 Memory cell in LSTM

By stacking multiple memory cells on top of each other, deep RNN can be created with the output sequence of one layer forming the input sequence for the next, which enable LSTM to discover information from a dynamically changing contextual window over the input sequence history rather than a static one as in the fixed-sized window applied in feed-forward neural networks [Sak et al., 2014].

6.7.2.2 Application of LSTM in anomaly identification

To deal with issues with high temporal dependency, a RNN is a natural choice due to the recurrent connections in the network, which allows the network to store memories of past information. However, standard RNN do not has the ability to learn long-term time dependencies because of the gradient vanishing problem as we discussed above. And LSTM can solve this fundamental problem by applying the special memory cells in the architecture [de Bruin et al., 2017]. By stacking memory cells, information of previous inputs x can be kept in the output to some degree, carried by cell state, which makes LSTM an outstanding tool to mimic time series. This is the reason why we would introduce this method for anomaly detection. The LSTM network leveraged in this experiment is constructed in python environment with Keras deep learning library running on top of TensorFlow library developed by Google. As mentioned above, each sampling units of raw vibration signals are divided into 10 parts before feature extraction. Therefore, the LSTM model is constructed to predict the 10th through the previous 9 parts. Each step includes five features in length. During the experiment, a selection of 500 samples is applied to train the LSTM neural network with 5-fold cross validation to validate the proposed approach. Figure 6.19 illustrates the numerical result of 5-fold cross validation, including the mean square errors of all the features and their average values.

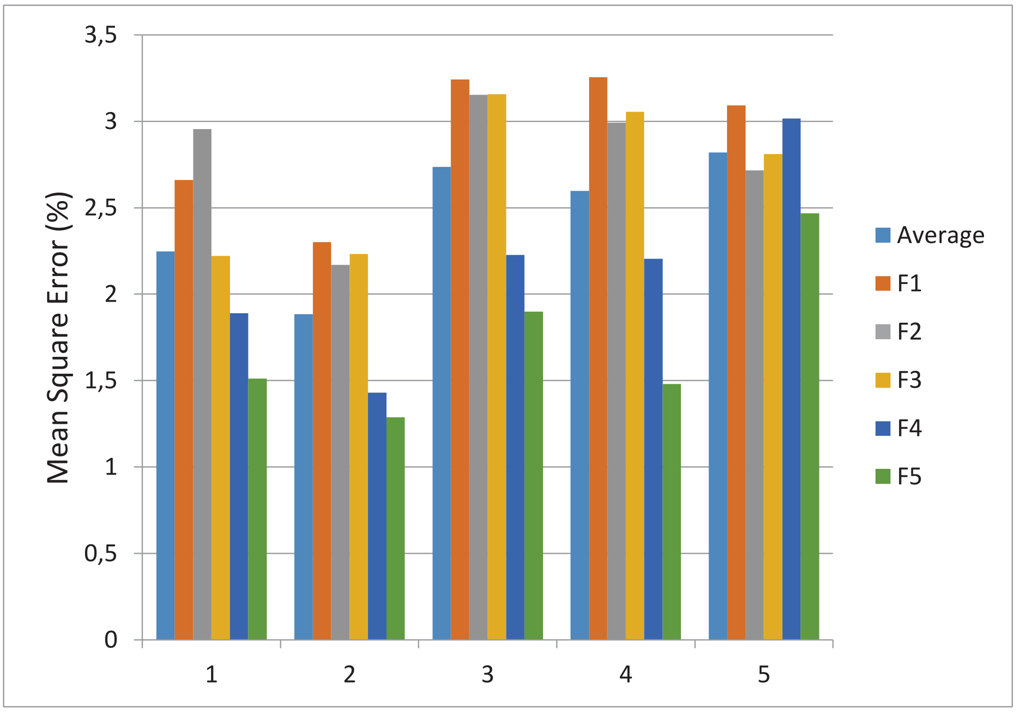

Figure 6.19 Numerical result of 5-fold cross validation

The numerical result of 5-fold cross validation shows that the proposed SAE-LSTM approach has the ability to predict multiple features sequence in certain degree. In addition, the stability of the method is also acceptable. Visually, the second run during the process of cross validation shares the best performance and will be leveraged to verify the performance of anomaly detection. Figure 6.20 and Figure 6.21 shows construction error with training epochs and test result of the constructed model during the training process, respectively. Visually, the mean construction error started to converge at about the 280th epochs with tiny fluctuation. The training errors of the LSTM neural network at all of the five feature sequences fluctuates between -0.4 to 0.3. The prediction result is not ideal. Since the target is to distinguish the anomaly and normal working condition instead of predicting the multiple features sequence directly. The performance of proposed method need to be further validated.

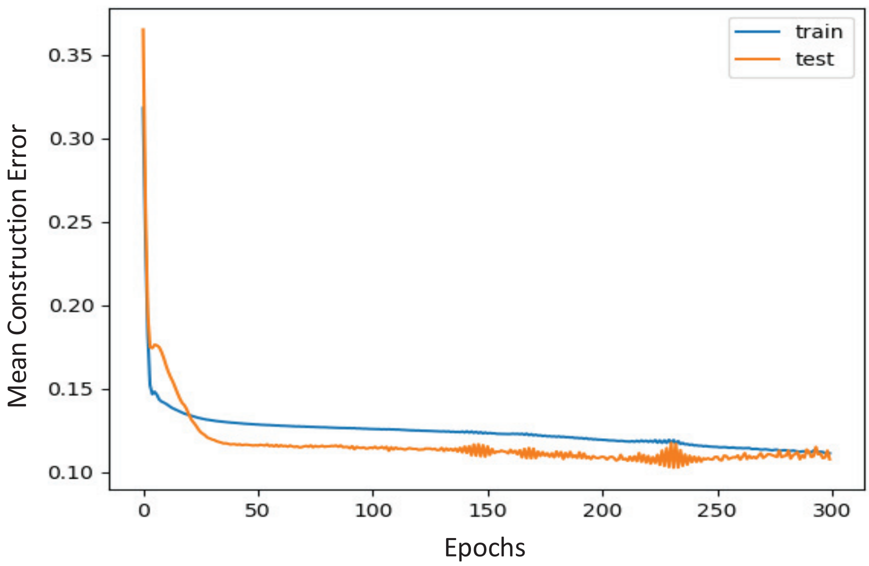

Figure 6.20 Construction error with training

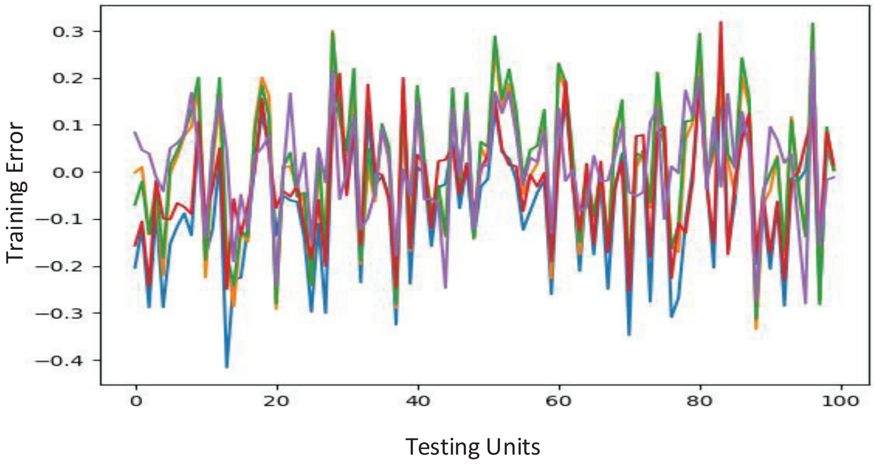

Figure 6.21 Training errors of five feature sequences

6.7.3 Validation and discussion

During the experiment, the multiple features sequence, which is obtained through SAE-based representation learning and unity-based normalization, is leveraged as the inputs of LSTM neural network. Figure 6.22 illustrates the process of proposed SAE-LSTM approach for anomaly detection. As mentioned above, the raw vibration signals was first divided into 10 parts. After SAE-based representation learning, the features at first 9 steps in each feature sequence will be used as inputs to map the features at 10th step during the training process. Therefore, the LSTM neural network is constructed with 9 LSTM memory cells to represent the previous 9 steps in multiple features sequence and predict the features at 10th step. The error between predicted and actual values of the features at 10th step will be leveraged to determine whether the equipment works in a normal condition.

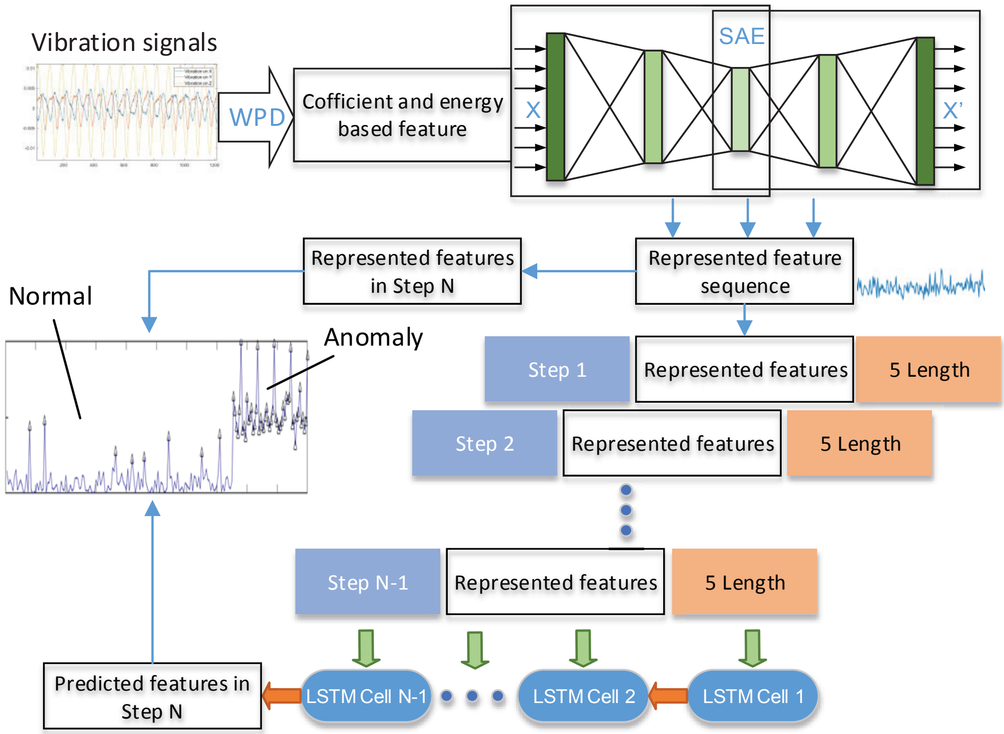

Figure 6.22 Process of SAE-LSTM approach for anomaly detection

To validate the performance of proposed SAE-LSTM approach for anomaly detection, a selection of 200 samples, constructed of 150 samples in normal condition and 50 sample in anomaly, will be leveraged for testing. Figure 6.23 shows the testing result of anomaly detection through SAE-LSTM.

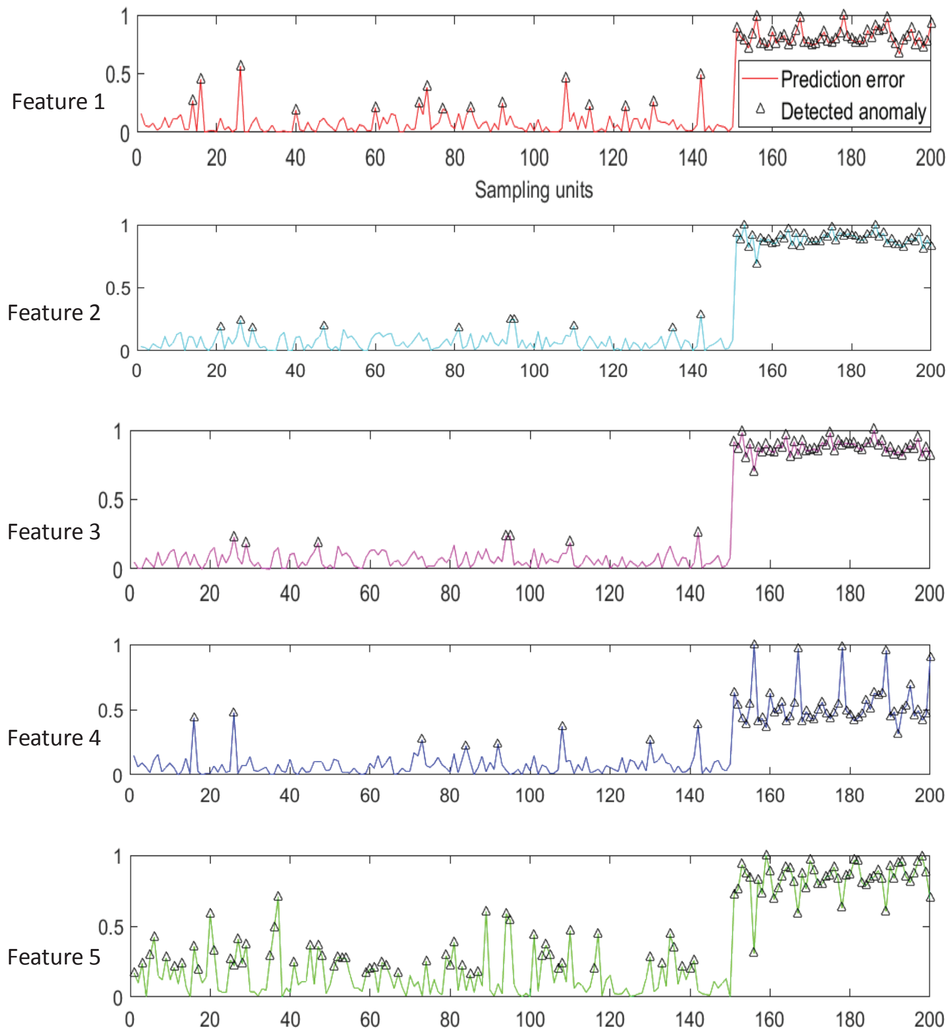

Figure 6.23 Result of anomaly detection through SAE-LSTM

During the test, we applied the largest MSE in each feature obtained through 5-fold cross validation as the criterion to detect anomaly in the equipment. Since the sensitivity of each feature to anomaly condition is highly subjective in nature. The criterion applied during the test is based on the overall performance in all features, which means only when all the prediction errors in five features are beyond the average values, the condition would be considered as anomaly. Table 6.5 listed the overall performance of proposed SAE-LSTM anomaly detection approach and the result of each single feature sequence, respectively.

Table 6.5 Performance of SAE-LSTM for anomaly detection

	Threshold (%)	Number of detected anomaly	Accuracy (%)
Feature 1	18.0418	65	92.5
Feature 2	17.7565	60	95
Feature 3	17.7676	57	96.5
Feature 4	17.3675	58	96
Feature 5	16.7939	109	71.5
F1&F2&F3&F4&F5		52	99

According to the numerical results, the performance and stability of proposed SAE-LSTM method for anomaly detection is acceptable. However, it is also obvious that not all the features trained through representation learning is suitable for anomaly detection. Since fault diagnosis is a subjective problem in nature, we guess that Feature 5 were trained to represent original features which is relatively irrelevant or insensitive to anomaly condition by SAE during the representation learning. In this research, the data-driven model was trained and validated in a completely unsupervised learning environment, which means the proposed SAE-LSTM approach could ideally detect anomaly working condition when the data is collected without labels. In practical applications, if part of the data is collected with labels, it may help to optimize the detection criterion and further improve the detection accuracy, which would be a direction for future research.

6.8 Summary

This chapter has introduced an experiment about the application of deep learning algorithms for fault classification, degradation assessment and anomaly detection in rotary machinery. In sensing, the art of anticipating failure and degradation in rotary machinery by means of monitoring vibration is the most efficient and widely applied in industry, since the measured vibration levels will change according to the defect or degradation of a rotary machine. Simultaneously, vibrations caused by the defects occur at specific vibration frequencies, characteristic of the components, their operation, assembly and wear. This is the reason why vibration signals are usually transformed into time domain, frequency domain, or time-frequency domain using analysing techniques such as FFT, STFT, WPD, or EMD. In this experiment, WPD is leveraged since it can efficiently represent information from vibration signals in both time and frequency domains. For decades, to construct a fault detection or degradation assessment system would require elaborate engineering and empirical knowledge in relevant fields to extract and select suitable features from the raw data as the perception to represent and interpret the signs of faults. During the experiment, all the original features represented through WPD are used as inputs without selection. Several data-driven models are used during the experiment to discover fault information and test their ability to detect, classify and estimate the degradation of failures. According to the numerical results, DNN demonstrate its unique superiority in degradation assessment for rotary machinery when the target condition is fully covered by historical data. A novel SAE-LSTM approach is also proposed for anomaly detection when all the data is collected without labels. The proposed approach provide an alternative method to leverage and integrate features for anomaly detection instead of empirical knowledge.

References

Asuhaimi Mohd Zin, A., Saini, M., Mustafa, M. W., Sultan, A. R., and Rahimuddin. (2015). New algorithm for detection and fault classification on parallel transmission line using DWT and BPNN based on Clarke's transformation. *Neurocomputing, 168*(Supplement C), 983-993. doi:https://doi.org/10.1016/j.neucom.2015.05.026

Baraldi, P., Cannarile, F., Di Maio, F., and Zio, E. (2016). Hierarchical k-nearest neighbours classification and binary differential evolution for fault diagnostics of automotive bearings operating under variable conditions. *Engineering Applications of Artificial Intelligence, 56*, 1-13.

Bengio, Y., Courville, A., and Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence, 35*(8), 1798-1828.

Bengio, Y., Lamblin, P., Popovici, D., and Larochelle, H. (2007). *Greedy layer-wise training of deep networks.* Paper presented at the Advances in neural information processing systems.

Candel, A., Parmar, V., LeDell, E., and Arora, A. (2015). Deep learning with h2o: H2O.

de Bruin, T., Verbert, K., and Babuška, R. (2017). Railway track circuit fault diagnosis using recurrent neural networks. *IEEE transactions on neural networks and learning systems, 28*(3), 523-533.

Deng, L. (2012). Three classes of deep learning architectures and their applications: a tutorial survey. *APSIPA transactions on signal and information processing*.

El Kadiri, S., Grabot, B., Thoben, K.-D., Hribernik, K., Emmanouilidis, C., von Cieminski, G., and Kiritsis, D. (2016). Current trends on ICT technologies for enterprise information systems. *Computers in Industry, 79*(Supplement C), 14-33. doi:https://doi.org/10.1016/j.compind.2015.06.008

Erhan, D., Bengio, Y., Courville, A., Manzagol, P.-A., Vincent, P., and Bengio, S. (2010). Why does unsupervised pre-training help deep learning? *Journal of Machine Learning Research, 11*(Feb), 625-660.

Ferreira, C. B. R., and Borges, D. b. L. (2003). Analysis of mammogram classification using a wavelet transform decomposition. *Pattern Recognition Letters, 24*(7), 973-982.

Galloway, G. S., Catterson, V. M., Fay, T., Robb, A., and Love, C. (2016). Diagnosis of tidal turbine vibration data through deep neural networks. *In: Proceedings of the Third European Conference of the Prognostics and Health Management Society 2016.*(PHM Society), 172-180.

Glorot, X., and Bengio, Y. (2010). *Understanding the difficulty of training deep feedforward neural networks.* Paper presented at the Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*: MIT press.

Goodfellow, I. J., Warde-Farley, D., Mirza, M., Courville, A., and Bengio, Y. (2013). Maxout networks. *arXiv preprint arXiv:1302.4389*.

Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., and Schmidhuber, J. (2016). LSTM: A search space odyssey. *IEEE transactions on neural networks and learning systems*.

Ha, D., Ahmed, U., Pyun, H., Lee, C.-J., Baek, K. H., and Han, C. (2017). Multi-mode operation of principal component analysis with k-nearest neighbor algorithm to monitor compressors for liquefied natural gas mixed refrigerant processes. *Computers & Chemical Engineering, 106*(Supplement C), 96-105. doi:https://doi.org/10.1016/j.compchemeng.2017.05.029

Hochreiter, S., and Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation, 9*(8), 1735-1780. doi:10.1162/neco.1997.9.8.1735

ISO 20816-1. (2016). *Mechanical vibration -- Measurement and evaluation of machine vibration*.

Jafar, R., Shahrour, I., and Juran, I. (2010). Application of Artificial Neural Networks (ANN) to model the failure of urban water mains. *Mathematical and Computer Modelling, 51*(9), 1170-1180.

Krauss, C., Do, X. A., and Huck, N. (2017). Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. *European journal of operational research, 259*(2), 689-702.

LeCun, Y. A., Bottou, L., Orr, G. B., and Müller, K.-R. (2012). Efficient backprop *Neural networks: Tricks of the trade* (pp. 9-48): Springer.

Lei, Y., Lin, J., He, Z., and Zuo, M. J. (2013). A review on empirical mode decomposition in fault diagnosis of rotating machinery. *Mechanical Systems and Signal Processing, 35*(1), 108-126.

Liao, L., and Ahn, H.-i. (2016). Combining Deep Learning and Survival Analysis for Asset Health Management. *International journal of prognostics and health management*.

Lin, J., and Chen, Q. (2014). A novel method for feature extraction using crossover characteristics of nonlinear data and its application to fault diagnosis of rotary machinery. *Mechanical Systems and Signal Processing, 48*(1), 174-187.

Lu, C., Wang, Z.-Y., Qin, W.-L., and Ma, J. (2017). Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification. *Signal Processing, 130*, 377-388.

Murugappan, M., Ramachandran, N., and Sazali, Y. (2010). Classification of human emotion from EEG using discrete wavelet transform. *Journal of Biomedical Science and Engineering, 3*(04), 390.

Nielsen, M. A. (2015). Neural networks and deep learning: Determination Press USA.

Okumura, S. (2011). *The short time Fourier transform and local signals.* Carnegie Mellon University.

Poultney, C., Chopra, S., and Cun, Y. L. (2007). *Efficient learning of sparse representations with an energy-based model.* Paper presented at the Advances in neural information processing systems.

Precup, R.-E., Angelov, P., Costa, B. S. J., and Sayed-Mouchaweh, M. (2015). An overview on fault diagnosis and nature-inspired optimal control of industrial process applications. *Computers in Industry, 74*(Supplement C), 75-94. doi:https://doi.org/10.1016/j.compind.2015.03.001

Ramesh Babu, N., and Jagan Mohan, B. (2017). Fault classification in power systems using EMD and SVM. *Ain Shams Engineering Journal, 8*(2), 103-111. doi:https://doi.org/10.1016/j.asej.2015.08.005

Recht, B., Re, C., Wright, S., and Niu, F. (2011). *Hogwild: A lock-free approach to parallelizing stochastic gradient descent.* Paper presented at the Advances in neural information processing systems.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1985). *Learning internal representations by error propagation*. Retrieved from

Sak, H., Senior, A., and Beaufays, F. (2014). *Long short-term memory recurrent neural network architectures for large scale acoustic modeling.* Paper presented at the Fifteenth Annual Conference of the International Speech Communication Association.

Scheffer, C., and Girdhar, P. (2004). *Practical machinery vibration analysis and predictive maintenance*: Elsevier.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks, 61*, 85-117.

Soualhi, A., Medjaher, K., and Zerhouni, N. (2015). Bearing health monitoring based on Hilbert–Huang transform, support vector machine, and regression. *IEEE transactions on instrumentation and measurement, 64*(1), 52-62.

Sun, Y.-j., Zhang, S., Miao, C.-x., and Li, J.-m. (2007). Improved BP neural network for transformer fault diagnosis. *Journal of China University of Mining and Technology, 17*(1), 138-142.

Sutskever, I., Martens, J., Dahl, G., and Hinton, G. (2013). *On the importance of initialization and momentum in deep learning.* Paper presented at the International conference on machine learning.

Ting, W., Guo-zheng, Y., Bang-hua, Y., and Hong, S. (2008). EEG feature extraction based on wavelet packet decomposition for brain computer interface. *Measurement, 41*(6), 618-625.

Vachtsevanos, G. J., Lewis, F., Hess, A., and Wu, B. (2006). *Intelligent fault diagnosis and prognosis for engineering systems*: Wiley Online Library.

Wang, P., and Guo, C. (2013). Based on the coal mine's essential safety management system of safety accident cause analysis. *American Journal of Environment, Energy and Power Research, 1*(3), 62-68.

Wang, Z.-Y., Lu, C., and Zhou, B. (2017). Fault diagnosis for rotary machinery with selective ensemble neural networks. *Mechanical Systems and Signal Processing*.

Xue, J.-Z., Zhang, H., Zheng, C.-X., and Yan, X.-G. (2003). *Wavelet packet transform for feature extraction of EEG during mental tasks.* Paper presented at 2003 international conference on the Machine learning and cybernetics.

Yang, D.-M., Stronach, A., MacConnell, P., and Penman, J. (2002). Third-order spectral techniques for the diagnosis of motor bearing condition using artificial neural networks. *Mechanical Systems and Signal Processing, 16*(2-3), 391-411.

Yang, Y., Dong, X., Peng, Z., Zhang, W., and Meng, G. (2015). Vibration signal analysis using parameterized time–frequency method for features extraction of varying-speed rotary machinery. *Journal of Sound and Vibration, 335*, 350-366.

Zhang, C., Lim, P., Qin, A., and Tan, K. C. (2017). Multiobjective deep belief networks ensemble for remaining useful life estimation in prognostics. *IEEE transactions on neural networks and learning systems*.

Zhang, Y., Liu, B., Ji, X., and Huang, D. (2016). Classification of EEG Signals Based on Autoregressive Model and Wavelet Packet Decomposition. *Neural Processing Letters*, 1-14.

Zhang, Z., and Zhao, J. (2017). A deep belief network based fault diagnosis model for complex chemical processes. *Computers & Chemical Engineering, 107*(Supplement C), 395-407. doi:https://doi.org/10.1016/j.compchemeng.2017.02.041

Zhao, R., Yan, R., Wang, J., and Mao, K. (2017). Learning to monitor machine health with convolutional bi-directional lstm networks. *Sensors, 17*(2), 273.

Chapter 7
Conclusion and future research

The work presented in this thesis is generally overviewed in this chapter and some suggestions for the future research are proposed.

7.1 Summary and conclusions

Industry 4.0 promotes the vision of smart manufacturing in future factories, where machines are connected as a collaborative community to share information, implement management and perform maintenance in a more reasonable and systematic way. Under this environment, predictive maintenance has attracted not only researchers' but also manufacturers' attention along with the development of data-driven methods since it is an ideal maintenance policy to minimize the cost of maintenance with the premise of zero failure manufacturing through the utilization of real-time data to forecast potential faults. However, there still exist some challenges and technology issues to implement predictive maintenance, which has been discussed in Chapter 3.

This thesis aims to bridge the gap between these fields and construct a framework for predictive maintenance concerning Industry 4.0 concept and industry big data, to monitor working condition of equipment, identify impending or potential failures, and minimize the number of unnecessary maintenance performance under the premise of zero failure manufacturing. The proposed framework consists of three tiers, namely data acquisition from multiple sources, fault identification and prediction, and decision support and maintenance implementation. The proposed framework can provide empirical knowledge for academics and practitioners to identify and prioritize their steps towards predictive maintenance and condition-based maintenance management under the environment of Industry 4.0.

In Chapter 2, a systematic investigation about deep learning approaches applied for fault identification and prediction are presented. Five types of deep learning approaches, which can improve or solve some 'bottlenecks' for predictive maintenance, are summarized along with the theoretical speculations about their superiorities for predictive maintenance owing to the special architectures. The thesis also provides evidence based on practical applications from literature and two experiments to interpret the superiorities of deep learning approaches in certain issues or with prerequisites for fault identification and prediction, which might offer a guidance to select the most suitable deep learning architecture for practical applications.

Chapter 4 provides helpful guidelines to formulate the steps for predictive maintenance with Industry 4.0 concepts in machining centers. The guidelines contain the entire process of fault analysis and treatment, which includes sensor and data acquisition, data preprocessing, fault diagnosis and prognosis, performance indicator analysis, and maintenance schedule optimization.

Chapter 5 demonstrates a case study of applying deep learning approach to predict backlash error for maintenance implementation scheduling in a machining center. A hierarchical diagnosis and prognosis system for backlash error detection and prediction based on DBN is proposed to deal with the situation when target condition is beyond the historical data. To provide a comprehensive comparison, two other intelligent algorithms, BPNN and SVMR, are also applied to replace the DBN as the prognosis model. The numerical results show the superiority to apply deep learning method for backlash error prediction. Moreover, a novel maintenance implementation strategy HDPS-BPSO is also proposed to illustrate the implementation of predictive maintenance in practical application. The numerical result also shows the benefit of implementing predictive maintenance compared with preventive one.

Chapter 6 introduced an experiment of deep learning algorithms for fault classification and degradation assessment in rotary machinery. The research focuses on the accuracy of impending failures identification and evaluation, which is the key to achieve predictive maintenance in many situations. During the experiment, WPD is leveraged to represent the information in both time and frequency domains from collected vibration signals. Several data-driven methods are used to detect, classify and estimate the degradation of failures. The numerical results demonstrate the superiority of DNN to evaluate the degradation for rotary machinery, which proves the theatrical speculation raised previously. A novel SAE-LSTM approach is also proposed for anomaly detection when all the data is collected without labels. The proposed methods provide alternative general approaches to leverage and integrate features for fault diagnosis instead of empirical knowledge.

7.2 Suggestion for future work

The following are proposed for future work:

At present, the structures of applied data-driven models for predictive maintenance are mainly selected based on empirical knowledge. It will be of interest to deduce a criterion or formulate certain guidelines to make this process automated.

Deep learning has demonstrated its ability to deal with the original represented information during the study. To test the performance of deep learning through the raw data directly may also be included in future work.

In this thesis, the focus mainly lies on the data mining process during the implementation of predictive maintenance. Research on maintenance management could also be developed in future work.